Appendix A (for Online Publication)

A-1 Model Setup

In this section, I outline an inclusive setup of the model that spans several assumptions, in both the nurse-managed system and the self-managed system. For detailed comments of the setup for the baseline model, refer to Section 3 in the main paper. Consider the following simple game, involving two physicians $j \in \{1, 2\}$ who work in a single pod at the same time:

1. The triage nurse (in the nurse-managed system) or physicians (in the self-managed system) commit to an assignment policy function if they are able to.

2. At time $t = 0$ both physicians receive one patient each, discovering $\theta_j \in \left\{\underline{\theta}, \overline{\theta}\right\}$, where $\overline{\theta} > \underline{\theta} > 0$. Type $\underline{\theta}$ occurs with probability p, and type $\overline{\theta}$ occur with probability $1 - p$.[1]

3. Physicians commit to how long they will keep their initial patients (t_j).

4. Physicians send a message $m_j \in \left\{\underline{\theta}, \overline{\theta}\right\}$ to the triage nurse (in the nurse-managed system) or to each other (in the self-managed system) if they are able to report their types. Otherwise they send $m_j = \emptyset$.

5. With probability $\psi > 0$, physicians observe each other's θ_j. Physician types are never observed by the triage nurse.

6. Exactly one patient will arrive at time t_a distributed with uniform probability across $\left[\underline{\theta}, \overline{\theta}\right]$. The physician who receives this third patient, denoted by$\mathcal{J}(3)$, is determined as follows:

    (a) In the nurse-managed system, the triage nurse assigns the patient based on physician censuses c_1 and c_2 (the number of patients they have, either 0 or 1, which is public information) at $t = t_a$. If she committed to an assignment policy function in Stage 1, she uses this. Otherwise, she decides assignment at t_a. If physicians can report their types, then her policy function can also use m_1 and m_2.

    (b) In the self-managed system, the physicians determine assignment.

        i. If physicians cannot commit to an assignment policy function, each physician independently decides whether to choose the new patient at each time $t \geq t_a$.

        ii. If physicians can commit to an assignment policy function, then the new patient is assigned according to this policy function at $t = t_a$. The policy function uses censuses c_1 and c_2, and observed types $o_j \in \{\theta_j, \emptyset\}$ ($o_j = \theta_j$ with probability ψ, $o_j = \emptyset$ with probability $1 - \psi$). If physicians can report their types, then the policy function also uses m_j.

[1] In this appendix, I will refer to θ_j as the type of physician j, since physicians start with one patient each and are otherwise identical.

7. Physicians complete their work on the one or two patients under their care and end their shifts. They receive the payoff

$$u_j^P = -\left(t_j - \theta_j\right)^2 - K_P\left(\theta_j\right)\mathbb{I}\left\{\mathcal{J}\left(3\right) = j\right\}, \tag{A-1.1}$$

where t_j is the time that physician j keeps his initial patient, $\theta_j \in \left\{\underline{\theta}, \overline{\theta}\right\}$ is the workload entailed by his patient and unobservable by the triage nurse, and $K_P\left(\theta_j\right) > 0$ is the cost of getting a potential third patient conditional on initial workload θ_j. I denote $K_P\left(\underline{\theta}\right) = \underline{K}_P$ and $K_P\left(\overline{\theta}\right) = \overline{K}_P$, and I impose that $\overline{K}_P > \underline{K}_P > 0$.

A-2 Nurse-managed System

In the nurse-managed system, the triage nurse assigns the new patient to a physician. I flexibly specify the triage nurse's utility as

$$u^N = -D \sum_{j \in \{1,2\}} \left(t_j - \theta_j\right)^2 - K_N\left(\theta_{j(3)}\right). \tag{A-2.1}$$

Notice the similarity between this utility function and that of the physicians, shown in (A-1.1). D is an indicator that allows the triage nurse to care about the treatment times of the first two patients as outcomes (if $D = 1$). Remember that the socially optimal discharge times for patients is $t_j = \theta_j$, which is universally agreed upon. The second term, $K_N\left(\theta\right)$, is the cost of assigning the new patient to a physician of type θ. I specify $K_N\left(\underline{\theta}\right) = 0$ and $K_N\left(\overline{\theta}\right) = \overline{K}_N$, and I impose $\overline{K}_N > 0$ but do not restrict the the value of $\overline{K}_N$ relative to $\overline{K}_P - \underline{K}_P$. This reflects that the triage nurse would like to assign the new patient to a physician with low workload.

The triage nurse's action is defined by an assignment policy function $\pi\left(c_1, c_2\right)$, where censuses $c_j \in \{0, 1\}$ are the numbers of patients the physicians have at time t_a.[2] To simplify notation, I impose that $\pi\left(0,0\right) = \pi\left(1,1\right) = \frac{1}{2}$ and $\pi \equiv \pi\left(0,1\right) = 1 - \pi\left(1,0\right)$. That is, when both physicians have equal censuses, assignment should not prefer to send the new patient to one physician or the other, since there is no other information about who is less busy at that time. Also, probabilities must sum to 1. To be clear, $\pi = 1$ represents *ex post* efficiency in that if one physician has no patients and the other has one, the former physician is known to be less busy with certainty.

In what follows, I will first show that the analysis is simplified by the fact that each physician's best response function is unaffected by what he expects the other physician to do. I then consider three different scenarios of whether physicians can report their types and of whether the triage nurse can commit to an assignment policy function. In my preferred model, I assume that physicians cannot report their types but that the triage nurse can credibly commit to an assignment policy. Although θ_j is a single index known with certainty in this model, in practice

[2] Below, I consider the scenario in which physicians can report their types to the triage nurse. In this case, the policy function takes the form $\pi\left(m_1, m_2\right)$, where $m_j \in \left\{\underline{\theta}, \overline{\theta}\right\}$.

workload is difficult to communicate, and ED physicians rarely report their workloads to the triage nurse.[3] On the other hand, what the triage nurse does when she observes c_1 and c_2 is easily observable; the ED management may even set guidelines which instruct her to assign not entirely by censuses, and more importantly, beliefs about her behavior can easily be updated in practice by physicians when there is more than one new patient arriving during their shifts. Finally, I slightly modify the model to communicate the intuition for identification strategy of using the expected flow of future patients as a driver of foot-dragging.

A-2.1 Irrelevance of Peer Strategy or Type

One advantage for the uniqueness of equilibria in this simple two-type model is that the physician's strategy does not depend on the strategy or type of his peer. To see this, consider some assignment function $\pi(c_1, c_2)$, which we will assume for now is based on censuses. Again, this assignment function can be summarized by a single parameter $\pi \equiv \pi(0,1)$. Any reasonable assignment rule sends the new patient to a less-busy physician, if there is one, with greater probability (i.e., $\pi > \frac{1}{2}$). It is easy to see that a high-type physician will then discharge his patient at $\bar{t} = \bar{\theta}$, because discharging his patient earlier only increases his changes of getting the new patient and discharging later has no benefit either.

Lemma A-1. *A high-type physician always discharges his patient at $\bar{t} = \bar{\theta}$.*

The next step is to solve for the best response function of the index physician who is of low type, given his peer's strategy. Denote the discharge time that a low-type peer will choose as $\underline{t}_{-j} \in \left[\underline{\theta}, \bar{\theta}\right]$. The expected utility of the index physician is

$$\mathbb{E}\left[u_j^P\left(\underline{t}_j; \pi, \underline{\theta}\right)\right] = -\left(\underline{t}_j - \underline{\theta}\right)^2 - \underline{K}_P \Pr\left\{\mathcal{J}(3) = j \left| \underline{t}_j, \pi\right.\right\},$$

The probability that $\mathcal{J}(3) = j$ can be written more explicitly as

$$\Pr\left\{\mathcal{J}(3) = j \left| \underline{t}_j, \pi\right.\right\} = \begin{cases} \frac{1}{2}\left(\frac{\underline{t}_j - \underline{\theta}}{\bar{\theta} - \underline{\theta}}\right) + \pi\left(\frac{\underline{t}_{-j} - \underline{t}_j}{\bar{\theta} - \underline{\theta}}\right) + \left(\frac{1}{2}p + \pi(1-p)\right)\left(\frac{\bar{\theta} - \underline{t}_{-j}}{\bar{\theta} - \underline{\theta}}\right), & \underline{t}_j < \underline{t}_{-j} \\ \begin{array}{c} \frac{1}{2}\left(\frac{\underline{t}_{-j} - \underline{\theta}}{\bar{\theta} - \underline{\theta}}\right) + \left((1-\pi)p + \frac{1}{2}(1-p)\right)\left(\frac{\underline{t}_j - \underline{t}_{-j}}{\bar{\theta} - \underline{\theta}}\right) + \\ \left(\frac{1}{2}p + \pi(1-p)\right)\left(\frac{\bar{\theta} - \underline{t}_j}{\bar{\theta} - \underline{\theta}}\right) \end{array}, & \underline{t}_j > \underline{t}_{-j} \end{cases}, \tag{A-2.2}$$

which is continuous at $\underline{t}_j = \underline{t}_{-j}$. This expression represents flow probabilities divided among three potential windows of time and uses the fact that the peer is low-type with probability p.

[3] Patient status in the ED is highly uncertain and multidimensional (e.g., patients expected to have the same length of stay may entail very different workloads due to severity). Physicians often do not share the same assessment when viewing the same patient, and information is invariably lost in communication, such that patient "hand-offs" are viewed as a patient safety issue (Apker et al., 2007). For similar reasons, it is highly unlikely that physicians may have a credible way report each other's workloads to the triage nurse, as conceived by Moore and Repullo (1988) as a subgame perfect mechanism. This is discussed more at the beginning of Section A-2.4.

Both $\underline{t}_{-j}$ and p additively affect expected utility, but they do not affect the first-order condition with respect to $\underline{t}_j$. In particular,

$$\frac{\partial}{\partial \underline{t}_j} \Pr\left\{\mathcal{J}(3) = j \left| \underline{t}_j, \pi\right.\right\} = \frac{1}{\overline{\theta} - \underline{\theta}} \left(\pi - \frac{1}{2}\right). \tag{A-2.3}$$

Increasing $\underline{t}_j$ always reduces the flow probability of receiving the third patient by $\pi - \frac{1}{2}$. Thus, his best response does not depend on the action of his (low-type) peer or the probability that his peer is low-type. It consequently is immaterial whether he knows his peer's type or the sequence in which physicians first choose t_j and then observe θ_{-j} with probability ψ.

Lemma A-2. *The best response of a low-type physician in the nurse-managed system does not depend on the action of his peer nor the probability that his peer is low-type.*

A-2.2 No Physician Reporting, No Triage Nurse Commitment

Without physician reporting or triage nurse commitment, in equilibrium, the triage nurse chooses the optimal assignment policy $\pi(c_1, c_2)$ at time t_a, given physician discharge strategies $\underline{t}^*$ and $\overline{t}^*$ for low- and high-type physicians, respectively, and given censuses c_1 and c_2. With no commitment, the triage nurse will always choose $\pi^* \equiv \pi^*(0,1) = 1$, the *ex post* efficient choice, and physicians will respond with discharge strategies $\underline{t}^*$ and $\overline{t}^*$.

Proposition A-3. *In the Perfect Bayesian Equilibrium in the nurse-managed system with no physician reporting and no triage nurse commitment, the triage nurse always assigns the new patient to the physician with census 0 when there is another physician with census 1, i.e.,* $\pi^* = 1$. *Low-type physicians will foot-drag, choosing* $\underline{t}^* > \underline{\theta}$, *as in Equation (A-2.4).*

The triage nurse's assignment policy is simple: She will assign the new patient to the physician with no patients if the other has one patient (i.e., $\pi^* = 1$). In this case, given any $\underline{t}^*$ and $\overline{t}^*$, she expects that the physician with $c_j = 0$ must be low-type whereas the one with $c_{-j} = 1$ must be high type. Otherwise, if $c_1 = c_2$, she cannot distinguish the two physicians and will assign the patient with equal probability to each.

In equilibrium, physicians will discharge their initial patients with this knowledge. As stated in Lemma A-1, high-type physicians will never want to mimic low-type physicians and will discharge their patients at time $t = \overline{\theta}$. On the other hand, low-type physicians will want to mimic high-type physicians at least temporarily by keeping their patients longer than socially optimal, since this reduces his likelihood of getting the new patient. (Low-type) physicians do not consider the type of their peer because of Lemma A-2. The first-order condition of a low-type physician's problem of $\max_{t_j} \mathbb{E}\left[u_j^P(t_j; \underline{\theta})\right]$ yields

$$\underline{t}^* = \underline{\theta} + \frac{K_P}{4\left(\overline{\theta} - \underline{\theta}\right)}. \tag{A-2.4}$$

This reflects the fact that there is always a first-order gain to reducing the likelihood of getting the new patient, compared to a second-order loss to prolonging the patient stay to more than socially optimal. That is, there will always be foot-dragging. In fact, for sufficiently high $\underline{K}_P$ or sufficiently low $\overline{\theta} - \underline{\theta}$, there may be full pooling in that all physicians will discharge their patient at $t = \overline{\theta}$.

In this subgame perfect equilibrium, the triage nurse correctly assigns patients to physicians who are less busy when she observes $c_1 \neq c_2$. But by doing so, she incentivizes physicians to foot-drag and actually reduces the probability of seeing $c_1 \neq c_2$ when $\theta_1 \neq \theta_2$. That is, while her assignment is *ex post* efficient, it is *ex ante* inefficient.

A-2.3 No Physician Reporting, Triage Nurse Commitment

I now allow for the triage nurse to commit to a policy function π_N. The equilibrium in this case is the same as in the previous case without commitment, except that the triage nurse chooses π_N^* at $t = 0$ and not at $t = t_a$.

Proposition A-4. *In the Perfect Bayesian Equilibrium in the nurse-managed system with no physician reporting but triage nurse commitment, the triage nurse assigns the new patient to the physician with census 0 when there is another physician with census 1 with some probability π_N^* as given by Equation (A-2.7), which lies between $\frac{1}{2}$ and 1. A low-type physician foot-drags weakly less than under no triage nurse commitment, discharging his patient at $\underline{t}^*$ that is earlier and closer to $\underline{\theta}$.*

To analyze this case, first note that the triage nurse will still never want to send the new patient with greater probability to a physician with $c_j > c_{-j}$. It therefore still holds that high-type physicians will never want to mimic low-type physicians, and that low-type physicians have some reason to mimic high-type ones. To see this, for any given π_N, the first-order condition for a low-type physician yields

$$\underline{t}^* = \underline{\theta} + \frac{\underline{K}_P}{2\left(\overline{\theta} - \underline{\theta}\right)} \left(\pi_N - \frac{1}{2}\right). \tag{A-2.5}$$

The first-order gain in temporary mimicry still exists relative to the second-order loss, as long as $\pi_N > \frac{1}{2}$.

When the triage nurse commits to an assignment policy function, she will choose π_N such that her expected utility is maximized. Her expected utility is

$$\begin{aligned} \mathbb{E}\left[u^N\left(\pi_N\right)\right] &= -Dp\left(\underline{t}^* - \underline{\theta}\right)^2 - (1-p)^2\,\overline{K}_N - \\ &\qquad 2p\,(1-p)\,\overline{K}_N \left[\frac{1}{2}\left(\frac{\underline{t}^* - \underline{\theta}}{\overline{\theta} - \underline{\theta}}\right) + (1-\pi_N)\left(\frac{\overline{\theta} - \underline{t}^*}{\overline{\theta} - \underline{\theta}}\right)\right]. \end{aligned} \tag{A-2.6}$$

Substituting (A-2.5) into (A-2.6) and solving the first-order condition yields the optimal assignment rule under commitment

$$\pi_N^* = \frac{1}{2} + \frac{4p(1-p)\overline{K}_N\left(\overline{\theta}-\underline{\theta}\right)^2}{\underline{K}_P\left(Dp\underline{K}_P + 4p(1-p)\overline{K}_N\right)}. \tag{A-2.7}$$

It is easy to see that π_N^* given by (A-2.7) can be less than 1, which implies that the triage nurse sometimes sends the new patient to the physician with census 0 even when there is another one with census 1. With $\pi_N^* < 1$, by (A-2.5), $\underline{t}^*$ is lower than when there is no triage nurse commitment.

The important general point is that for some parameters the triage nurse will commit to an assignment policy function $\pi_N^* < 1$. Even if she only cares about the assignment of the third patient, as in Equation (A-2.8), she will commit to an *ex post* inefficient assignment policy in order to improve *ex ante* assignment. Under triage nurse commitment to $\pi_N^* < 1$, the degree of foot-dragging by low-type physicians, in Equation (A-2.5), will be lower than under no commitment, stated in Equation (A-2.4). The intuition for this is similar to the finding in Milgrom and Roberts (1988) that managers will sometimes choose to ignore valuable but distortable information from "influence activities."

This result is even stronger when the triage nurse cares about length of stay for the initial patients (i.e., $D = 1$). To see this, consider the optimal policy function if the triage nurse does not care about lengths of stay for the first two patients as outcomes and only cares about the assignment of the third patient (i.e., $D = 0$):

$$\pi_{N|D=0}^* = \frac{1}{2} + \frac{\left(\overline{\theta}-\underline{\theta}\right)^2}{\underline{K}_P}. \tag{A-2.8}$$

The triage nurse's choice of $\pi_{N|D=0}^*$ only depends on the *low-type physician's* cost of getting the new patient, because it is this physician that will engage in foot-dragging and distort information. In the case with $D = 1$, in (A-2.7), $\pi_{N|D=1}^*$ also increases as $\overline{K}_N$ or as p is closer to $\frac{1}{2}$. Since she cares about foot-dragging as an outcome, she will increase $\pi_{N|D=1}^*$ as she cares more about inefficient assignment (as $\overline{K}_N$ is higher) or as it is more likely she will encounter two physicians with different censuses (p is closer to $\frac{1}{2}$). Importantly, comparing (A-2.8) and (A-2.7), $\pi_{N|D=0}^* > \pi_{N|D=1}^*$. Also caring about foot-dragging as an outcome lowers her choice of π_N^* to reduce foot-dragging further.

A-2.4 Physician Reporting, Triage Nurse Commitment

I next consider the case in which physicians can also report their types as $m_j \in \left\{\underline{\theta}, \overline{\theta}\right\}$, which is a mechanism design problem without transfers. In the equilibrium in this case, the triage nurse provides a menu $\left\{t\left(\underline{\theta}\right), t\left(\overline{\theta}\right); \pi\left(m_1, m_2\right)\right\}$ to physicians, subject to an incentive compatibility constraint for physicians to report $m_j = \theta_j$ truthfully. Because the triage nurse immediately distributes the new patient upon arrival, there is an additional "intertemporal" constraint in that a physician who lies and then later reveals that he is lying can still be punished sufficiently so that he will not want to engage in this strategy. Given this menu, physicians will tell the truth.[4]

I find that, in contrast to physician signaling (no reporting to the triage nurse), there is no foot-dragging under the mechanism of physician reporting when there are discrete types. However, this is not the case with continuous types, in which physicians always have a first-order incentive to foot-drag. I show results for the case of continuous types in Section A-4, but the intuition is consistent with that of Myerson and Satterthwaite (1983): Efficiency is improved somewhat artificially when one restricts the types (and messages) that physicians can report. The restriction of the type space is quite severe in the ED (and most medical contexts), as patient care and physician workload are multidimensional and contain quite a lot of information, to the extent that communication problems exist even between doctors with no incentive for moral hazard (e.g., Apker et al., 2007). Multidimensional screening is even more complex than screening along a single continuous dimension and has been discussed in Rochet and Stole (2003).

Proposition A-5. *In the Perfect Bayesian Equilibrium in the nurse-managed system with physician reporting and triage nurse commitment, the triage nurse implements truth-telling by setting the assignment policy π_R^* equal to the assignment policy implied by intertemporal incentive compatibility, π_{IT}^*, given in Equation (A-2.12). Given truth-telling and discrete types, there is no foot-dragging, i.e., $\underline{t}^* = \underline{\theta}$. Assignment is still* ex post *inefficient with $\pi_R^* < 1$, but* ex ante *assignment efficiency is improved relative to no physician reporting.*

By Lemma A-1, there is no incentive compatibility constraint for high-type physicians, but there is one for low-type physicians. Specifically, the utility for a low-type physician under truth-telling must be at least as great as his utility if he were to report that he is a high-type physician, where I again use the notation $\pi \equiv \pi\left(\underline{\theta}, \overline{\theta}\right)$ to summarize any assignment policy:

$$-\left(t\left(\overline{\theta}\right)-\underline{\theta}\right)^2-\frac{1}{2}\underline{K}_P \leq -\pi_{IC}\underline{K}_P. \tag{A-2.9}$$

Using the fact that $t\left(\overline{\theta}\right) = \overline{\theta}$ by Lemma A-1,[5] this incentive compatibility constraint is actually binding at the optimal triage nurse assignment when she only cares about foot-dragging as a signal, stated in Equation (A-2.8):

$$\pi_{IC}^* = \frac{1}{2} + \frac{\left(\overline{\theta}-\underline{\theta}\right)^2}{\underline{K}_P}. \tag{A-2.10}$$

Note that although π_{IC}^* equals $\pi_{N|D=0}^*$ in Equation (A-2.8), the intuitions for the two expressions are different. In the case of π_{IC}^*, the constraint sets utility to be the same for a low-type physician

[4] Again, expectations over the peer's type are conveniently ignored due to Lemma A-2.

[5] In some mechanism design problems, this value of a high-type agent could be distorted upwards. However, because I have assumed that there is 0 probability that the new patient will arrive after $t = \overline{\theta}$, this mechanism cannot be enforced.

under truth-telling and under reporting to be high-type. In the case of $\pi^*_{N|D=0}$, triage nurse utility happens to be maximized at an assignment policy that causes a low-type physician to foot-drag midway with $\underline{t}^* = \underline{\theta} + \left(\overline{\theta} - \underline{\theta}\right)/2$. Also recall that $\pi^*_{N|D=0} > \pi^*_{N|D=1}$; so $\pi^*_{IC} \leq \pi^*_{N}$.

However, there is a second "intertemporal" incentive compatibility constraint that results from the facts that reports of θ_j and discharge times t_j are intertemporally separated and that the triage nurse is limited in her ability to punish a physician who reports $\theta_j = \overline{\theta}$ but discharges his patients before $t\left(\overline{\theta}\right) = \overline{\theta}$. That is, a low-type physician may report that he is high-type but discharge his patient at $\hat{t} < \overline{\theta}$. Without punishment (i.e., if he simply reverts to having the truth-telling flow probability of $\pi_{IT}\left(\underline{\theta}, \overline{\theta}\right)$ after t) he would be strictly better off by this scheme. To prevent this, there needs to be punishment, but the highest flow probability that the triage nurse can use from that point on is 1.

The second constraint is thus

$$-\left(\hat{t} - \underline{\theta}\right)^2 - \underline{K}_P \left[\frac{1}{2}\left(\frac{\hat{t} - \underline{\theta}}{\overline{\theta} - \underline{\theta}}\right) + \frac{\overline{\theta} - \hat{t}}{\overline{\theta} - \underline{\theta}}\right] \leq -\pi_{IT}\underline{K}_P. \quad \text{(A-2.11)}$$

This can be addressed by first solving for the optimal "cheating" $\hat{t}^*$ under full punishment:

$$\hat{t}^* = \underline{\theta} + \frac{\underline{K}_P}{4\left(\overline{\theta} - \underline{\theta}\right)},$$

which of course is the same as the optimal discharge time with no physician reporting or triage nurse commitment, stated in (A-2.4). This time can then be substituted into (A-2.11) in order to state the constraint on π_{IT}:

$$\pi^*_{IT} = 1 - \frac{\underline{K}_P}{16\left(\overline{\theta} - \underline{\theta}\right)^2}. \quad \text{(A-2.12)}$$

Note that $\pi^*_{IT} < 1$ for $\underline{K}_P > 0$. The intutition for this is that, with intertemporal incentive compatibility, the triage nurse can never implement $\pi_R = 1$, because if she did, then a low-type physician will always be better off by reporting to be high-type for some time. Revealing that he lied entails "punishment" which cannot be worse than the $\pi_R = 1$ he would have gotten with truth-telling anyway.[6]

It can be shown that $\pi^*_{IC} \leq \pi^*_{IT}$. In particular, $\pi^*_{IC} < \pi^*_{IT}$, for $\underline{K}^P / \left(\overline{\theta} - \underline{\theta}\right)^2 \in (0, 4)$, when there is a temptation to foot-drag but when $\underline{t}^* < \overline{\theta}$ in the case without physician reporting or triage nurse commitment, as in Section A-2.3 and Equation (A-2.4). The intuition is that it is weakly more difficult to implement the intertemporal constraint than the standard incentive compatibility constraint because it is always at least as easy for low-type physicians to "temporarily"

[6]In contrast, $\pi^*_{IC} = 1$ for $\underline{K}^P / \left(\overline{\theta} - \underline{\theta}\right)^2 \leq 2$. In this parameter space, a low-type physician is better off by telling the truth than by "fully" lying by reporting that he is high-type and keeping his patient until $t\left(\overline{\theta}\right) = \overline{\theta}$. Similarly, \pi_{N}^{*} $= 1$ for $\underline{K}^P / \left(\overline{\theta} - \underline{\theta}\right)^2 \leq 2$, because the temptation to foot-drag is sufficiently low so that the triage nurse is better off by incurring the maximum foot-dragging and the highest *ex post* assignment efficiency.

lie rather than "fully" lie. Thus, the intertemporal constraint will always be binding.

The assignment policy under physician reporting is therefore $\pi_R^* = \min(\pi_{IC}^*, \pi_{IT}^*) = \pi_{IT}^*$. Again, the assignment policy under no physician reporting but triage nurse commitment is equivalent to the assignment policy implied by the constraint with "full" lying: $\pi_{N|D=0}^* = \pi_{IC}^*$. At first glance, this suggests that assignment is less efficient *ex post* under physician reporting than under no reporting: $\pi_{N|D=0}^* < \pi_R^*$. This may not hold if the triage nurse also cares about foot-dragging on the initial patients ($D = 1$) under no reporting, since $\pi_{N|D=0}^* > \pi_{N|D=1}^*$.

However, the more important point is that under reporting and truth-telling, *ex post* assignment efficiency is also *ex ante* assignment efficiency, since physicians do not foot-drag. Further, the triage nurse's *ex ante* utility is strictly greater with physician reporting than with no reporting. To formalize this, I consider $\mathbb{E}\left[u^N\right]$, where $u^N$is given in Equation (A-2.1), which simplifies to $\overline{K}^N \Pr\left\{\theta_{j(3)} = \overline{\theta}\right\}$ when $D = 0$. Figure A-2.2 shows the difference in this value between physician reporting and no reporting, normalizing $\overline{K}^N = 1$. Note that this efficiency gain is strictly greater when $D = 1$, because the triage nurse also cares about foot-dragging on the initial patients as negative outcomes, of which there is none under reporting.

In summary, Proposition A-5 derives from the triage nurse's access to better information from physician reports. Subject to maintaining truth-telling, she can implement an *ex ante* assignment policy more efficient than the one without physician reporting, when she must balance *ex post* assignment efficiency with distortion of signals by physicians. In both cases, she is limited by the ability of physicians to distort signals or misreport the truth. In this sense, there is also a parallel intuition between the Milgrom and Roberts (1988) prediction and the standard mechanism design feature of information rents.

In this two-type model, there is no foot-dragging because the triage nurse uses π_R^* in order to implement truth-telling. There is no point in using $t(\underline{\theta})$ to implement truth-telling, because doing so would only make a low-type physician worse off under truth-telling, and $t\left(\overline{\theta}\right)$ is irrelevant because of the intertemporal incentive compatibility constraint. However, I will show in Section A-4 that this does not hold for continuous types, because with local incentive compatibility constraints, the benefit of foot-dragging ("full" lying in the mechanism design framework) at the truth ($t_j = \theta_j$) is first-order while its cost is second-order. As argued above, this is an important concern for the ED setting (and most other medical settings) in which information is quite complex, and in which the restriction to discrete types is highly artificial.

A-2.5 Flow of Expected Future Work

This simple model assumes a single patient will arrive in the interval $t \in \left[\underline{\theta}, \overline{\theta}\right]$, which is convenient for capturing the temptation for moral hazard by low-type physicians. However, there are of course in practice usually many more new patients, and my main identification for foot-dragging will be the response of lengths of stay to the flow of expected future work. One way to capture the intuition of expected future work is to modify the model so that a new patient is expected to arrive in the interval $t \in [\underline{\theta}, \underline{\theta} + \Delta t]$, maintaining the assumption of a single new patient. I also allow Δt to be greater or less than $\overline{\theta} - \underline{\theta}$, although I assume that physicians may only be assigned the new patient prior to $t = \overline{\theta}$ for $\Delta t > \overline{\theta} - \underline{\theta}$ and focus on interior solutions for $\Delta t < \overline{\theta} - \underline{\theta}$.[7]

I again focus on the behavior of low-type physicians, who has an incentive to foot-drag and mimic high-type physicians. It is easy to see that all denominators in fractions in Equation (A-2.2) should now be Δt instead of $\overline{\theta} - \underline{\theta}$, and $\overline{\theta}$ should be replaced by $\underline{\theta} + \Delta t$. Equation (A-2.3) should then have Δt in the denominator rather than $\overline{\theta} - \underline{\theta}$, at least for interior $\underline{t}_j \in \left[\underline{\theta}, \overline{\theta} + \Delta t\right]$:

$$\frac{\partial}{\partial \underline{t}_j} \Pr \left\{ j(3) = j \left| \underline{t}_j \in \left[\underline{\theta}, \overline{\theta} + \Delta t\right], \pi \right. \right\} = \frac{1}{\Delta t} \left(\pi - \frac{1}{2} \right).$$

Respective denominators in the foot-dragging Equations (A-2.4) and (A-2.5) should also now have Δt instead of $\overline{\theta} - \underline{\theta}$ for interior solutions $\underline{t}^* \in \left[\underline{\theta}, \overline{\theta} + \Delta t\right]$. For example, in the baseline scenario of no physician reporting but triage nurse commitment, the equilibrium foot-dragging by low-type physicians is

$$\underline{t}^* = \underline{\theta} + \frac{K_P}{2\Delta t} \left(\pi - \frac{1}{2} \right).$$

This slight modification communicates the intuition that as the expected future work increases (decreases), the marginal temptation to foot-drag increases (decreases) because the certainty of receiving a new patient within each infinitessimal unit of time around discharge is greater (smaller).

A-3 Self-managed System

In this section, I will analyze foot-dragging and patient assignment in the self-managed system. I assume the same physician utilities and information structure (i.e., that they observe each other's θ_j with probability $\psi > 0$) as I did for the nurse-managed system. The only difference will be that the two physicians, not a triage nurse, are responsible for deciding who gets the new patient. In what follows, I will also consider analagous cases in which physicians may or may not report θ_j to each other and in which they may or may not be able to commit to an assignment policy function based on censuses. The key difference between results for all of these cases and corresponding results for the nurse-managed system derives from physicians *both* observing each other's θ_j with probability $\psi > 0$ *and* being able to use that information in patient assignment.

The assignment of patients by the physicians themselves deserves further mention. In the case in which physicians are unable to commit to an assignment policy, I represent assignment as a non-cooperative bargaining game in which physicians can choose to see the new patient. At any time after the new patient has arrived, as long as the patient remains unchosen, either

[7] I maintain the assumption of a single new patient, so that I do not have to consider capacity constraints and physician strategy for subsequent patients, in order to keep the model simple. I also focus on interior solutions so that this single patient is relevant for comparative statics.

physician may choose to see the new patient or wait. If one physician chooses the patient, he gets the patient with probability 1. If they both choose the patient at the same time, they each get the patient with probability $\frac{1}{2}$. This game is very much motivated by Rubinstein's (1982) non-cooperative bargaining game in which two players with complete information about each other's costs bargain over a good that declines in value over time. Aided by a setup that is simpler than Rubinstein's, I will also extend its analysis to consider incomplete information about peer types.

In the other case in which physicians can commit to an assignment policy, I represent an assignment policy function that is quite similar to the one I defined previously for the nurse-managed system. Here the probability of assignment to physician 1 is a function of censuses and potential observations of type $o_j \in \{\emptyset, \theta_j\}$, where recall that both types are observed with probability ψ. That is, the policy function takes the form $\pi_S(c_1, c_2, o_1, o_2)$. It is easy to see that physicians should commit to $\pi_S\left(c_1, c_2, \underline{\theta}, \overline{\theta}\right) = 1$ for all c_1 and c_2. Similar to before, the one relevant parameter to be solved for is $\pi_S \equiv \pi_S(0, 1, \emptyset, \emptyset)$. Commitment in this case involves physicians jointly determining an assignment rule to maximize expected utility prior to knowing their types.

A-3.1 No Physician Reporting, No Policy Commitment

I first consider the case in which physicians can neither report their types nor commit to an assignment policy. In order to allow a war of attrition, I need to elaborate on the cost of waiting to treat a patient who has arrived. While I consider this extended physician utility, which accounts for the cost of delay in assignment, it will be obvious that this utility function is equivalent to the baseline utility function (A-1.1) used in the rest of the conceptual framework (and in the main text) when there is no delay.

A-3.1.1 Extended Physician Utility

Denote τ as the time that has elapsed since the new patient arrived. Assume that both physicians incur a cost of having the new patient remain untreated at each time τ, increasing over τ, and also assume now that the cost of getting the new patient also increases over τ but not as quickly. I then extend physician utility as

$$u_j^P = -\left(t_j - \theta_j\right)^2 - W_j\left(\theta_j; \tau^*, j(3)\right) - K_P\left(\theta_j\right) \mathbb{I}\{j(3) = j\}. \tag{A-3.1}$$

The new second term $W_j(\cdot)$ is a generalized cost of waiting. Denoting τ^* as the elapsed time that it took for the new patient to be chosen by someone, and $j(3)$ as the physician who chose the patient, W_j is simply integrated over each τ.

$$W_j\left(\theta_j; \tau^*, j(3)\right) = \int_0^{\tau^*} \omega_j\left(\tau; \theta_j, \tau^*, j(3)\right) d\tau,$$

$\omega_j(\tau;\theta_j,\tau^*,j(3))$ is a flow cost that depends on τ and whether physician j received the patient at τ or not:

$$\omega_j(\tau;\theta_j,\tau^*,j(3)) = \begin{cases} w\tau, & \text{if } \tau \neq \tau^* \\ k_0(\theta_j) + k\tau, & \text{if } \tau = \tau^*,\ j = j(3) \\ 0, & \text{if } \tau = \tau^*,\ j \neq j(3). \end{cases}$$

$w\tau$ is the flow cost of waiting imposed on both physicians as long as the patient remains unchosen. $k_0(\theta_j)+k\tau$ is the initial flow cost of seeing the patient, where I denote $\underline{k}_0 = k_0(\underline{\theta})$ and $\overline{k}_0 = k_0(\overline{\theta})$ for brevity and assume $\underline{k}_0 < \overline{k}_0$. Note that since $\tau = \tau^*$ with mass of 0, I can simplify the cost-of-waiting integral to $W_j(\tau^*) = \frac{1}{2}w\tau^{*2}$.

I assume that the flow costs of seeing that patient, starting with the initial cost $k_0(\theta_j)$ if the patient is seen immediately, integrate to $K_P(\theta_j)$. This ensures that the extended utility in (A-3.1) reduces to Equation (A-1.1) when there is no delay in assignment.[8] To ensure single crossing, I impose that k and the continued flow cost of seeing the new patient are both less than w. Finally I assume that physicians are responsible for taking care of patients up to $t = \overline{\theta} + \overline{k}_0/(w-k)$ to ensure that any patient arriving up to $t = \overline{\theta}$ could be chosen by a physician.

A-3.1.2 Non-cooperative Assignment by Physicians

In order to analyze assignment by physician choice, I consider four different cases. The first two cases occur when physicians observe each other's types, which happens with probability ψ. Case 1 is that physicians observe that one is type $\underline{\theta}$ and the other is type $\overline{\theta}$ and proceeds quite similarly to Rubinstein's (1982) setup with complete information. Physicians will infer that a low-type physician will weakly prefer choosing the patient at $\underline{\tau} = \underline{k}_0/(w-k)$, while a high-type physician will weakly prefer choosing the patient at $\overline{\tau} = \overline{k}_0/(w-k)$ and will strictly prefer waiting at $\underline{\tau} < \overline{\tau}$. Therefore, by the subgame where physicians choose to see the patient at $\underline{\tau}$, a low-type physician will be assigned the patient with probability 1. Given that there is a cost to waiting, a low-type physician will prefer to see the patient immediately than to wait until $\underline{\tau}$. So in equilibrium, a low-type physician will choose the patient at $\tau^* = 0$ with probability 1.

Case 2 is that physicians observe that they are both the same type $\theta \in \{\underline{\theta},\overline{\theta}\}$. There exists no pure strategy Nash equilibrium here. To see this, if there exists a Nash equilibrium, under symmetry both physicians should choose the patient at some $\tau^*(\theta)$. But one physician would always be strictly better off by waiting and letting the other physician choose the patient. There does exist a mixed strategy Nash equilibrium though. I generally denote the probability with which each physician will choose the patient at each elapsed time τ as $q(\theta;\theta,\tau)$, where the first argument is the type θ_j of the index physician (in this case shared) and the second argument

[8] This is automatic in the nurse-managed system and assumed in the case of physician commitment to a policy function in the self-managed system. I could also consider that the triage nurse might delay the assignment of patients to physicians if she can gain more information about their types. However, this would not add any useful intuition for the nurse-managed system, and of course, the setup of physicians starting with one patient each is itself an abstraction.

denotes the observation $o_{-j} \in \{\theta_{-j}, \varnothing\}$ of the peer's type (in this case observed). To satisfy the conditions for indifference between choosing and not choosing,

$$q(\theta;\theta,\tau) = \frac{\max\{0, w\tau - (k_0(\theta) + k\tau)\}}{w\tau - (k_0(\theta) + k\tau)/2}, \tag{A-3.2}$$

From Equation (A-3.2), note that $q(\theta,\tau) > 0$ only when the flow cost of waiting is inefficiently greater than the flow cost of initially treating the patient, i.e., $\tau > k_0(\theta)/(w-k)$. Also, although $q(\theta,\tau)$ increases with τ, $\lim_{\tau\to\infty} q(\theta) = (w-k)/(w-k/2) < 1$.

The next two cases occur in the subgame, occurring with probability $1-\psi$, in which physicians do not observe each other's types. As in the nurse-managed system, a high-type physician will still discharge his patient at $t = \overline{\theta}$. Suppose that a low-type physician discharges his patient at $t = \underline{t}^*$. Types can then be perfectly deduced only during times $t \geq \underline{t}^*$. So prior to $\underline{t}^*$, we have a game of incomplete information, and if the new patient arrives at $t_a < \underline{t}^*$, physicians will have an incentive to wait because types are unknown and therefore certain times for patient choice (i.e., Case 1) are impossible. Thus, during any $t \in T_{\varnothing}[t_a, \underline{t}^*]$, physicians must decide whether to choose patients while peer types are unknown. Case 3 considers the choice equilibrium strategy for a low-type physician during this time. As in Case 2, he will purely wait if $\tau \leq \underline{k}_0/(w-k)$ and will engage in a mixed strategy if $\tau > \underline{k}_0/(w-k)$. Similar to Equation (A-3.2),

$$q(\underline{\theta};\varnothing,\tau) = \min\left\{1, \frac{\max\{0, w\tau - (\underline{k}_0 + k\tau)\}}{p(w\tau - (\underline{k}_0 + k\tau)/2)}\right\}. \tag{A-3.3}$$

Equation (A-3.3) shows that a low-type physician is actually more likely to choose the new patient when he does not know that his peer's type relative to when he knows his peer is also low-type (but obviously less likely than when he knows his peer is high-type). Also, if the probability p of having a low-type peer is low enough (i.e., if $p < (w-k)/(w-k/2)$), he may even choose the patient with certainty at some point.

Case 4 considers the strategy for a high-type physician during the same $T_{\varnothing} = [t_a, \underline{t}^*]$. For a given elapsed time since arrival τ, note that both high-type and low-type physicians cannot be engaging in a mixed strategy. This would require

$$\frac{w\tau - (\underline{k}_0 + k\tau)}{w\tau - (\underline{k}_0 + k\tau)/2} = pq(\underline{\theta};\varnothing,\tau) + (1-p)\,q(\overline{\theta};\varnothing,\tau) = \frac{w\tau - (\overline{k}_0 + k\tau)}{w\tau - (\overline{k}_0 + k\tau)/2},$$

which is impossible for finite τ, given $\underline{k}_0 < \overline{k}_0$. So in equilibrium, high-type physicians will wait until after the time when low-type physicians should have chosen the new patient with certainty, which is implied in (A-3.3) to be

$$\underline{\tau}^* = \frac{\left(1 - \frac{1}{2}p\right)\underline{k}_0}{(1-p)\,w - \left(1 - \frac{1}{2}p\right)k}.$$

Note that types will then be revealed at $t_R = \min\left(\underline{t}^*, t_a + \underline{\tau}^*\right)$.

Finally, note that if the patient is still unchosen at t_R when physician types are revealed, then by definition, both physicians must be high type, and they will know this. They should mix with probability as defined in (A-3.2) as in Case 2. They could wait until some time after t_R until they start mixing, or they could immediately start mixing at t_R, depending on whether $t_R - t_a > \overline{k}_0 / (w - k)$. As in Case 2, each physician will never choose the patient with certainty.

A-3.1.3 No Incentive for Foot-dragging

Considering all 4 cases in the section above, it is clear that low-type physicians can never hope to have the new patient assigned to a high-type peer. This is true even when types are unknown by peers, because as shown in Case 4, high-type physicians will still wait until any low-type physician would have chosen the patient with certainty before choosing the patient with any positive probability.

Another point that is obvious from the above analysis of physician assignment without policy commitment is that physicians will often delay choosing patients, if they know that they are equally busy or if they are unsure how busy their peer is. In fact, under this delay, patients may sometimes never be chosen, given that there is no commitment to choose patients eventually. The delay in choosing patients represents the channel of "free-riding." However, this represents an *ex ante* utility loss at the stage that low-type physicians decide $\underline{t}^*$, since physicians only start placing some positive probability on choosing the new patient when they would have chosen the patient anyway if there were no peer.

Proposition A-6. *In the Perfect Bayesian Equilibrium in the self-managed system with no physician reporting or commitment to an assignment policy, there will be no foot-dragging, i..e,* $\underline{t}^* = \underline{\theta}$. *If the patient is chosen and if there is a low-type physician, assignment will always be to a low-type physician. However, physicians will wait to choose the new patient (free-ride) if they are of the same type.*

Low-type physicians have no incentive to conceal their types by foot-dragging for two reasons: First, low-type physicians can never hope to have a potential high-type peer choose the new patient before them. Second, concealing that they are low-type (foot-dragging) only leads to free-riding, which represents an *ex ante* utility loss. As in the nurse-managed system, assignment is completely *ex post* efficient in the sense that patients are never assigned to physicians with lower censuses when censuses differ. In addition, given that there is no foot-dragging, this also implies *ex ante* efficiency. However, there is a new "assignment" inefficiency of free-riding in that physicians may delay seeing patients and sometimes not even get to see them despite preferring to had there been no peer. In this model, free-riding only occurs when physicians are of the same type, since types can always be inferred by the time the new patient arrives at $t \in \left[\underline{\theta}, \overline{\theta}\right]$, as low-type physicians never foot-drag.[9]

[9] In practice, types may not be perfectly deduced (which would also happen in the model if the new patient

A-3.1.4 Remark on Continuous Types

I will conclude this subsection with a remark on continuous types in order to show that with continuous types the intuition for the stark result in Proposition A-6 does not hold. To see this, first note that with continuous types, even if types are unknown, the probability that a physician has the same type as his peer has mass 0. Thus, physicians should choose patients with pure strategies in equilibrium.

When physician do not observe each other's types (with probability $1-\psi$), physicians then use peer signals to infer θ_{-j}, and by subgame perfection physician j will choose the new patient at $\tau = 0$ if and only if $c_j > c_{-j}$. We now have a similar situation as in the traditional system with no triage nurse commitment, in Section A-2.3, where there is a first-order gain to foot-dragging. So in equilibrium, there should be no free-riding but positive foot-dragging. However, recall that with probability ψ, physicians observe each other's types. In this case, the physician j will choose the new patient at $\tau = 0$ if and only if $\theta_j > \theta_{-j}$, regardless of c_j and c_{-j}. Because of this, physicians will foot-drag less than they would have under the traditional system with no triage nurse commitment.

A-3.2 No Physician Reporting, Policy Commitment

From the analysis in Section A-3.1, it is clear that without physician commitment to an assignment policy, there could be large welfare losses in the form of free-riding and patients going untreated. This suggests scope for improvement by committing to an assignment policy. As a practical rationale, physicians often divide work before they can deduce each other's true workloads if new patients need to be seen in a timely manner.

Introduced above, the policy function takes the form $\pi_S(c_1, c_2, o_1, o_2)$, where o_j is type of physician j if observed and null otherwise. The following are obvious: $\pi_S\left(c_1, c_2, \underline{\theta}, \overline{\theta}\right) = 1$ for all c_1 and c_2; $\pi_S(c_1, c_2, \theta_1, \theta_2) = \frac{1}{2}$ for $\theta_1 = \theta_2$ and all c_1 and c_2; and $\pi_S(c_1, c_2, \emptyset, \emptyset) = \frac{1}{2}$ for $c_1 = c_2$.[10] As in the nurse-managed system, the assignment policy can then be represented by a single parameter $\pi_S \equiv \pi_S(0, 1, \emptyset, \emptyset)$. In equilibrium, physicians choose the optimal assignment policy π_S^* at time $t = 0$, given physician discharge strategies $\underline{t}^*$ and $\overline{t}^*$ for low- and high-type physicians, respectively. Given this assignment policy, physicians choose the optimal discharge strategies $\underline{t}^*$ and $\overline{t}^*$.

Proposition A-7. *In the Perfect Bayesian Equilibrium in the self-managed system with no physician reporting but with commitment to an assignment policy, if as $\psi > 0$, $\Delta K_P > \overline{K}_N$, and $D = 1$, then there will be less foot-dragging and more* ex post *efficient assignment than in the nurse-managed system with no physician reporting but triage nurse commitment.*

---

could arrive at $t < \underline{\theta}$), and this could support free-riding. However, as shown in the main paper, free-riding does not appear to be significant empirically, which suggests that physicians can commit to an assignment policy or have sufficient information about each other's types (either by censuses or observations of true workload).

[10]For simplicity and for consistency with the nurse-managed system, I assume that patients are immediately assigned under this policy. See also footnote (8).

The scenario for no physician reporting and policy commitment is analyzed similarly as in the corresponding scenario in the nurse-managed system. For any given policy π_S, a low-type physician will discharge his patient at time

$$\underline{t}^* = \underline{\theta} + (1-\psi)\frac{\underline{K}_P}{2\left(\overline{\theta}-\underline{\theta}\right)}\left(\pi_S - \frac{1}{2}\right). \tag{A-3.4}$$

Note again the similarity between Equations (A-2.5) and (A-3.4). The only difference is that the second term is multiplied by $1-\psi$, because with probability ψ, foot-dragging will have no effect on assignment. Self-management – both the observation of true workload and the use of this information in assignment – decreases foot-dragging relative to the nurse-managed system.

It now follows that the physicians will choose a policy function that takes Equation (A-3.4) into consideration in order to maximize their *ex ante* utilities, before they have received their initial patients. They expected to be type $\underline{\theta}$ with probability p and type $\overline{\theta}$ with probability $1-p$, and they maximize

$$\begin{aligned} \mathbb{E}\left[u^P\left(\pi_S\right)\right] \quad = \quad & -p\left(\underline{t}^*-\underline{\theta}\right)^2-(1-p)^2\,\Delta K_P - \\ & 2p\,(1-p)\,(1-\psi)\,\Delta K_P\left[\frac{1}{2}\left(\frac{\underline{t}^*-\underline{\theta}}{\overline{\theta}-\underline{\theta}}\right)+(1-\pi_S)\left(\frac{\overline{\theta}-\underline{t}^*}{\overline{\theta}-\underline{\theta}}\right)\right], \end{aligned} \tag{A-3.5}$$

where I conveniently transform the expected utility by subtracting $\underline{K}_P$ and using $\Delta K_P \equiv \overline{K}_P - \underline{K}_P$. Again, note the similarity with Equation (A-2.6), with two differences. First, with probability ψ, the policy function is irrelevant for assignment, so $1-\psi$ appears in the third term.[11] Second, instead of $\overline{K}_N$, the analogous parameter ΔK_P from the physician's utility function is used because physicians are the ones making patient assignment.

Ex ante, physicians would like to avoid assigning the new patient to a busier physician because it is more costly, by ΔK_P, for that physician to deal with that patient. However, *ex post*, once physicians know their types, low-type physicians have the moral hazard to avoid the new patient. Commitment to a policy function allows physicians *ex ante* to balance their desire for proper assignment with the knowledge that proper assignment will cause costly moral hazard.

Maximizing (A-3.5) with respect to π_S, after substituting (A-3.4) for $\underline{t}^*$, yields the optimal policy function

$$\pi_S^* = \frac{1}{2} + \frac{4p\,(1-p)\,\Delta K_P\left(\overline{\theta}-\underline{\theta}\right)^2}{(1-\psi)\,\underline{K}_P\left(p\underline{K}_P + 4p\,(1-p)\,\Delta K_P\right)}. \tag{A-3.6}$$

The differences between (A-2.7) and (A-3.6) are twofold. First, the denominator is multiplied by $1-\psi$, which reflects the fact that foot-dragging is lessened by the possible observation of true workload, and which improves the efficiency of assignment even when true workload is

[11] The third term also represents the efficiency loss with respect to misassignment under any policy commitment. It is larger with small ψ, p close to $1/2$, and large ΔK_P. On the other hand, the efficiency loss with no policy commitment is larger with large k, large w, or p close to 0 or 1. Depending on these parameters, physicians may opt for no policy function even if they can commit to one.

not observed. Second, ΔK_P replaces $\overline{K}_N$. Thus, as long as $\psi > 0$ and $\Delta K_P > \overline{K}_N$, then $\pi^*_{N|D=1} < \pi^*_S$.

Although I cannot definitively say that one is bigger than the other, it is likely that $\Delta K_P > \overline{K}_N$. To see this, notice that the triage nurse's utility function in Equation (A-2.1) can include the outcomes of *both* physicians. $\overline{K}_N$ represents the amount that she values assignment of the third patient compared to the amount that she values the average of outcomes for the first two patients. On the other hand, in the physician utility function in Equation (A-1.1), the cost of receiving another patient is scaled relative to the outcome of a single patient. So if the triage nurse had similar preferences as both physicians, we should have $\Delta K_P \approx 2\overline{K}_N$. For both of these reasons, π^*_S again should be greater than π^*_N, meaning that the self-managed system improves the efficiency of assignment relative to the nurse-managed system.

A-3.3 Physician Reporting, Policy Commitment

Finally, in the case of physician reporting and policy commitment, it suffices to consider how the observation of true workloads between peers and its use in the policy function modifies incentive compatibility constraints for a low-type physician.

The standard incentive compatibility constraint, assuming "full" lying, is

$$-\left(t\left(\overline{\theta}\right)-\underline{\theta}\right)^2-\frac{1}{2}\left(1-\psi\right)\underline{K}_P \leq -\pi_S\left(1-\psi\right)\underline{K}_P. \tag{A-3.7}$$

Even if he mimics a high-type physician, he will still be observed as a low-type physician with probability ψ, in which case mimicry was useless. This relaxes the incentive compatibility constraint that was originally (A-2.9) and allows a higher π_S to support truth-telling.

However, as in the nurse-managed system, the intertemporal incentive compatibility constraint will be binding. Recall that this constraint, shown in Equation (A-2.11) for the nurse-managed system, derives from the concern that low-type physicians can lie about their type and then discharge their patient earlier than $t\left(\overline{\theta}\right)=\overline{\theta}$. The constraint in the self-managed system is

$$-\left(\hat{t}-\underline{\theta}\right)^2-\left(1-\psi\right)\underline{K}_P\left[\frac{1}{2}\left(\frac{\hat{t}-\underline{\theta}}{\overline{\theta}-\underline{\theta}}\right)+\frac{\overline{\theta}-\hat{t}}{\overline{\theta}-\underline{\theta}}\right] \leq -\pi_S\left(1-\psi\right)\underline{K}_P, \tag{A-3.8}$$

where $\hat{t}$ is the time that a low-type physician reveals that he was lying when he initially reported that he was high-type. Again, with probability ψ, this lie will not pay off, which relaxes the incentive compatibility constraint to allow a higher π_S. Thus, assignment will be more efficient in the self-managed system compared to the nurse-managed system in this case as well.

Proposition A-8. *In the Perfect Bayesian Equilibrium in the self-managed system with physician reporting and commitment to an assignment policy, assignment will be more* ex ante *(and* ex post*) than in the nurse-managed system with physician reporting and triage nurse commitment. There is still no foot-dragging, given truth-telling and discrete types.*

A-4 Continuous Types

In this section, I relax the baseline two-type model to consider a continuum of types in the nurse-managed system with physician reporting. The purpose of this extended model is to communicate the intuition that, in contrast to the two-type analysis in Section A-2.4 with physician reporting, there will still be foot-dragging and inefficient assignment, as long as types are sufficiently rich. As discussed in Section A-2.4, this is much more reflective of reality, particular in settings in which information is complex, the very source of physician discretion in the first place. While I restrict attention to the nurse-managed system, similar intuition follows for the self-managed system.

The model is identical to the model outlined in Section A-1 except for two differences: First, each of the two physicians can be of type $\theta_j \in \left[\underline{\theta}, \overline{\theta}\right]$, drawn from some distribution which I do not specify.[12] Second, because types lie in a continuum, I allow for a more flexible triage nurse policy function that takes the form of $\pi\left(\theta_1, \theta_2, t\right)$, which is the flow probability of the physician 1 receiving a patient who arrives at time t when the types of physicians 1 and 2 are θ_1 and θ_2, respectively. Of course, this policy function cannot be reduced to a single parameter, because both θ_1 and θ_2 are continuous. In addition, I allow for the fact that the optimal policy may have a non-constant flow-rate for a given θ_1 and θ_2. The sufficient statistic from the standpoint of physician and triage nurse utility is

$$P\left(\theta_1, \theta_2\right) \equiv \frac{1}{\overline{\theta} - \underline{\theta}} \int_{\underline{\theta}}^{\overline{\theta}} \pi\left(\theta_1, \theta_2, t\right) dt,$$

which is the cumulative probability over time that physician 1 will receive the new patient, given that patient arrival is uniformly distributed and that assignment is immediate upon arrival. The reason I allow for the flexible time-dependent flow is to specifically consider "intertemporal feasibility" constraints in which $\pi\left(\theta_1, \theta_2, t\right) \leq 1$, for all θ_1, θ_2, t, which I have noted in Section A-2.4 as feature different from the standard screening problem.

I will analyze this model as follows: First, I will formalize the intertemporal feasibility constraints as incentive compatibility constraints for truth-telling. That is, physicians should have no incentive to discharge their patient at time $t < t\left(\hat{\theta}_j\right)$, where $\hat{\theta}_j$ is their reported type, which may be different than their true type θ_j. Second, as in the standard mechanism design problem, I will show that I can restrict attention to local incentive compatibility constraints in which physicians have no incentive to report the type continuously adjacent to theirs, conditional on the previous requirement that they follow the discharge time required by that report. Third, I will show by perturbation arguments that the optimal triage nurse policy function is continuous and strictly increasing. This implies that there will be positive foot-dragging and *ex post* inefficient

[12] The distribution is not important for this analysis, which shows positive foot-dragging and inefficient assignment, but it would be necessary to consider for an analysis that computes the optimal assignment function in closed form.

assignment in the sense that $P(\theta_j, \theta_{-j}) < 1$ for any $\theta_j < \theta_{-j}$. That is, in the remainder of this section, I will show the following:

Proposition A-9. *In the Perfect Bayesian Equilibrium in the nurse-managed system with physician reporting and triage nurse commitment, for a continuum of physician types distributed along* $[\underline{\theta}, \overline{\theta}]$, *there will be positive foot-dragging such that* $t(\theta) > \theta$ *for all* $\theta < \overline{\theta}$ *and* ex post *inefficient assignment such that* $P(\theta_j, \theta_{-j}) < 1$ *for any* $\theta_j < \theta_{-j}$.

A-4.1 Intertemporal Feasibility

As in any truth-telling equilibrium, physicians must not have the incentive to misreport their types. This setting in particular requires me to address the possibility of a physician reporting $\hat{\theta}_j$, in order to get some flow probability $\pi\left(\hat{\theta}_j, \theta_{-j}, t\right)$ of assignment for the new patient, but discharging the current patient at some time $t_j < t\left(\hat{\theta}_j\right)$. If physicians can receive a lower $P\left(\hat{\theta}_j, \theta_{-j}\right)$ by reporting $\hat{\theta}_j > \theta_j$ but not have to keep their patient as long as would be required by $t\left(\hat{\theta}_j\right)$, then they could be strictly better off. The reason for this departure from the standard screening model is that the cumulative probability $P(\theta_1, \theta_2)$ is not given by the triage nurse in a lump sum but rather over time. Thus, there is an "intertemporal feasibility" constraint in that the triage nurse may not be able to take back (in terms of lower probability of assignment) what she has already given in the past. More precisely, this constraint derives from the fact that even punishment policy functions are limited by $\pi(\theta_1, \theta_2, t) \leq 1$, for all θ_1, θ_2, and t.

In order to account for this, I simply consider that, for the standard mechanism design problem to work here, I need physicians to have no incentive to report $\hat{\theta}_j$ and then discharge their patient at time $t_j < t\left(\hat{\theta}_j\right)$ as opposed to reporting $\tilde{\theta}_j$ and discharging their patient at the same time $t_j = t\left(\tilde{\theta}_j\right)$. That is, if the physician is planning to discharge his patient at some time t_j, he may as well report the type that corresponds to that time. Note that this incentive compatibility does not yet require that the physician prefers to report $\tilde{\theta}_j = \theta_j$, which I consider later. In order to sustain this aspect of truth-telling, I assume that if the triage nurse catches a physician lying *and* deviating by $t_j < t\left(\hat{\theta}_j\right)$, then she will punish him by assigning him the new patient with probability 1 – but no more by the feasibility constraint – if the new patient arrives at $t \in \left[t_j, \overline{\theta}\right]$.

The incentive compatibility constraint implied by intertemporal feasibility then can be stated as

$$\begin{aligned} P\left(\tilde{\theta}_j, \theta_{-j}\right) &\equiv \frac{1}{\overline{\theta} - \underline{\theta}} \int_{\underline{\theta}}^{\overline{\theta}} \pi\left(\tilde{\theta}_j, \theta_{-j}, t\right) dt \\ &\leq \frac{1}{\overline{\theta} - \underline{\theta}} \left[\int_{\underline{\theta}}^{t\left(\tilde{\theta}_j\right)} \pi\left(\hat{\theta}_j, \theta_{-j}, t\right) dt + \overline{\theta} - t\left(\tilde{\theta}_j\right)\right], \forall \hat{\theta}_j < \tilde{\theta}_j, \theta_{-j} \end{aligned}$$

where I ignore the potential cost of discharging a patient at time $t\left(\tilde{\theta}_j\right)$, since this is held constant. This incentive compatibility constraint can equivalently be stated as

$$\int_{t(\theta_j)}^{\underline{\theta}} 1-\pi\left(\theta_j,\theta_{-j},t\right)dt \geq \int_{\underline{\theta}}^{t(\theta_j)} \pi\left(\theta_j,\theta_{-j},t\right)-\pi\left(\hat{\theta}_j,\theta_{-j},t\right)dt,\ \forall\hat{\theta}_j<\theta_j,\theta_{-j}. \tag{A-4.1}$$

This formalizes the intuition that the triage nurse requires scope for punishment in the policy function in order to prevent physicians from misreporting their types.

A-4.2 Local Incentive Compatibility

The remainder of the analysis proceeds similarly to the standard screening mechanism design. The triage nurse offers a menu of choices $\{(t(\theta_1,\theta_2),P(\theta_1,\theta_2))\}$ to physicians. Physicians simultaneously report θ_1 and θ_2, and by the Revelation Principle, they report truthfully.[13] The triage nurse's problem is to choose a menu that maximizes her expected utility subject to physician truth-telling.

I first show that the single-crossing condition holds. Physician utility remains the same as in the baseline model, but I account for continuous types by allowing the cost incurred by being assigned the new patient, $K_P(\theta)$, to be a continuous and differentiable function such that $K_P^{'}(\theta)\geq 0$. It is easy to show that

$$\frac{\partial}{\partial\theta_j}\left[-\frac{\partial u_j^P/\partial P\left(\theta_j,\theta_{-j}\right)}{\partial u_j^P/\partial t_j}\right]>0$$

as long as $t>\theta$ and $K_P^{'}(\theta)\geq 0$. The former condition that $t>\theta$, equivalent to foot-dragging, will be shown later to hold in equilibrium. I will assume the latter more strictly by $K_P^{'}(\theta)>0$.

Given the single-crossing condition, the set of incentive compatibility constraints,

$$-\left(t\left(\theta_j,\theta_{-j}\right)-\theta_j\right)^2-K_P\left(\theta_j\right)P\left(\theta_j,\theta_{-j}\right)\geq-\left(t\left(\hat{\theta}_j,\theta_{-j}\right)-\theta_j\right)^2-K_P\left(\theta_j\right)P\left(\hat{\theta}_j,\theta_{-j}\right),$$

for all $\theta_j,\theta_{-j},\hat{\theta}_j$, can be summarized as a monotonicity condition and local incentive compatibility constraints. The monotonicity condition is

$$\frac{\partial P\left(\theta_j,\theta_{-j}\right)}{\partial\theta_j}\leq 0, \tag{A-4.2}$$

and the local incentive compatibility constraints are

$$-2\left(t\left(\theta_j,\theta_{-j}\right)-\theta_j\right)\frac{\partial t\left(\theta_j,\theta_{-j}\right)}{\partial\theta_j}=K_P^{'}\left(\theta_j\right)\frac{\partial P\left(\theta_j,\theta_{-j}\right)}{\partial\theta_j}, \tag{A-4.3}$$

[13] In the following analysis, I assume that physicians know each other's types, but the intuition follows if they only act according to expected values of their peer's type.

simply stated by setting the derivative of the physician utility function equal to 0.[14]

A-4.3 Optimization Problem

I will now analyze the triage nurse's optimization problem subject to (A-4.1) and (A-4.3). Rather than solve her problem in closed form, I can obtain my results – that there will be foot-dragging and *ex post* inefficient assignment in equilibrium – by simple perturbation arguments.

I first maintain the assumption that the triage nurse's utility function takes the form as stated in (A-2.1), and I similarly assume that $K_N(\theta)$ is continuous and twice differentiable and that furthermore $K_N^{'}(\theta) > 0$. It then follows that she will offer menu options $t(\theta_j, \theta_{-j})$ and $P(\theta_j, \theta_{-j})$ that are continuous and differentiable in θ_j for all θ_j and θ_{-j}. To see this, suppose that her optimal assignment policy function $P(\theta_j, \theta_{-j})$ is discontinuous at $\theta_j = \theta_1$ and some $\theta_{-j} = \theta_2$, such that $\lim_{\varepsilon \to 0} P(\theta_1 - \varepsilon, \theta_2) - P(\theta_1 + \varepsilon, \theta_2) = \Delta > 0$. If this is the case, then no physician whose type $\theta_j \in (\theta_1 - \epsilon, \theta_1)$, for some $\epsilon > 0$, will truthfully reveal his type. The triage nurse cannot maintain truth-telling over some interval of types of strictly positive measure and therefore cannot implement her policy function, which is a contradiction.

Next, note that it is never optimal to have $t(\theta) < \theta$, as it reduces the discharge time away from what is socially optimal and only reduces the utility of the physician when truth-telling, which can only make the truth-telling constraint more costly. Given this, I can show that the optimal assignment policy $P(\theta_j, \theta_{-j})$ is strictly decreasing in θ_j. To see this, suppose that that the optimal policy is such that $P(\theta_j + \varepsilon, \theta_{-j}) \geq P(\theta_j, \theta_{-j})$ for some small $\varepsilon > 0$ and for some θ_j, θ_{-j}. However, the triage nurse can strictly increase her utility and maintain truth-telling by increasing $P(\theta_j, \theta_{-j})$ by some small $\epsilon > 0$ while keeping her set of $\{t(\theta_j, \theta_{-j})\}$ unchanged if $\partial t(\theta_j, \theta_{-j}) / \partial \theta_j > 0$, or by increasing $P(\theta_j, \theta_{-j})$ by some small $\epsilon > 0$ and concurrently decreasing $t(\theta_j, \theta_{-j})$ by some small $\delta > 0$ if $\partial t(\theta_j, \theta_{-j}) / \partial \theta_j \leq 0$. Note that this satisfies the monotonicity condition in (A-4.2).

Given that $P(\theta_j, \theta_{-j})$ is strictly decreasing in θ_j (by contradiction), $t(\theta) \geq \theta$ (by contradiction), and $K_P^{'}(\theta_j) > 0$ (by assumption), then (A-4.3) implies that $\partial t(\theta_j, \theta_{-j}) / \partial \theta_j > 0$ and $t(\theta_j) > \theta_j$, except for type $\theta_j = \overline{\theta}$, which need not be bound by an incentive compatibility constraint. That is, discharge times increase with the physician's type, regardless of his peer's type, and there is positive foot-dragging in equilibrium for all types $\theta_j < \overline{\theta}$.

Finally, *ex post* inefficient assignment derives from two independent facts shown above, either of which is sufficient. First, the continuity of $P(\theta_j, \theta_{-j})$ guarantees *ex post* inefficient assignment because it is impossible to have

$$P(\theta_j, \theta_{-j}) = \begin{cases} 1, & \theta_j < \theta_{-j} \\ 0, & \theta_j > \theta_{-j} \end{cases}, \qquad \text{(A-4.4)}$$

[14] An additional condition for the local incentive compatibility constraint to be valid is that the menu options $t(\theta_j, \theta_{-j})$ and $P(\theta_j, \theta_{-j})$ are continuous and differentiable in θ_j for all θ_j and θ_{-j}. This will also be shown below.

for all θ_j, θ_{-j}, without a discontinuity at $\theta_j = \theta_{-j}$. Second, inefficient assignment can also be proven by using intertemporal feasibility alone, as stated in (A-4.1). This constraint implies that it is impossible to have both $P(\theta_j, \theta_{-j}) > P\left(\hat{\theta}_j, \theta_{-j}\right)$ and $P(\theta_j, \theta_{-j}) = 1$, for any $\hat{\theta}_j$, θ_j, and θ_{-j}, and therefore also rules out (A-4.4).

References

Apker, Julie, Larry A. Mallak, and Scott C. Gibson, "Communicating in the Gray Zone: Perceptions about Emergency Physician-hospitalist Handoffs and Patient Safety," *Academic Emergency Medicine*, October 2007, *14* (10), 884–894.

Milgrom, Paul and John Roberts, "An Economic Approach to Influence Activities in Organizations," *The American Journal of Sociology*, 1988, *94* (Supplement), 154–179.

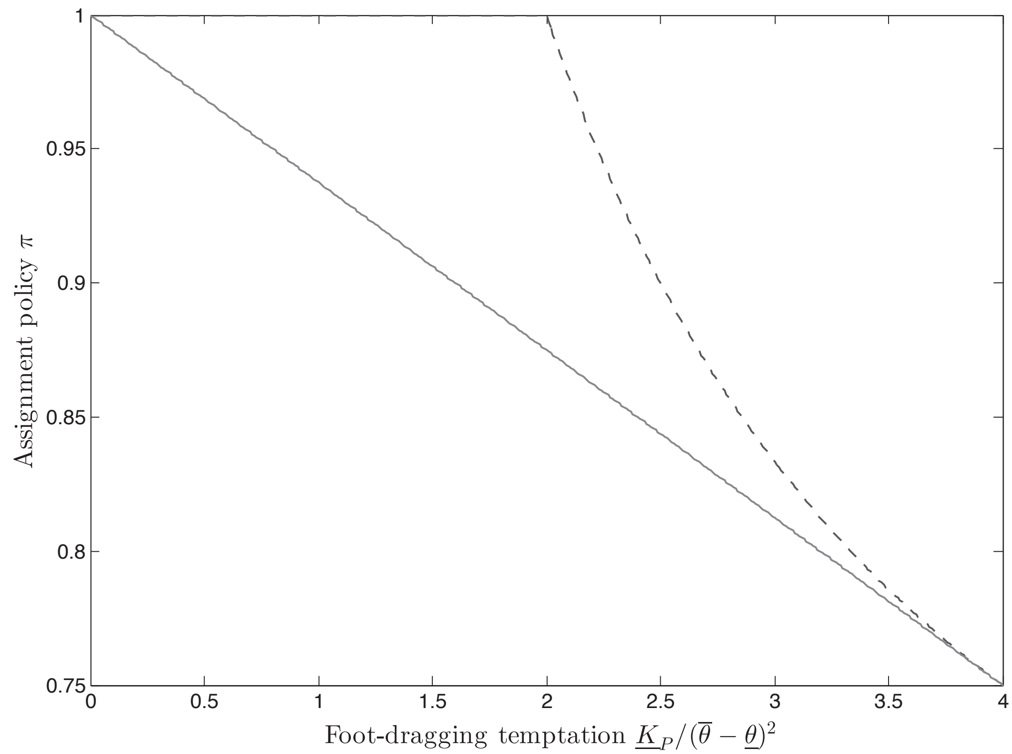
Moore, J. and R. Repullo, "Subgame perfect implementation," *Econometrica*, 1988, *56* (5), 1191–1220.

Myerson, Roger B and Mark A Satterthwaite, "Efficient mechanisms for bilateral trading," *Journal of Economic Theory*, April 1983, *29* (2), 265–281.

Rochet, Jean-Charles and Lars Stole, "The Economics of Multidimensional Screening," in M. Dewatripont, L. Hansen, and S. Turnovsky, eds., *Advances in Economics and Econometrics*, Vol. 35, Cambridge University Press, January 2003, pp. 150–197.

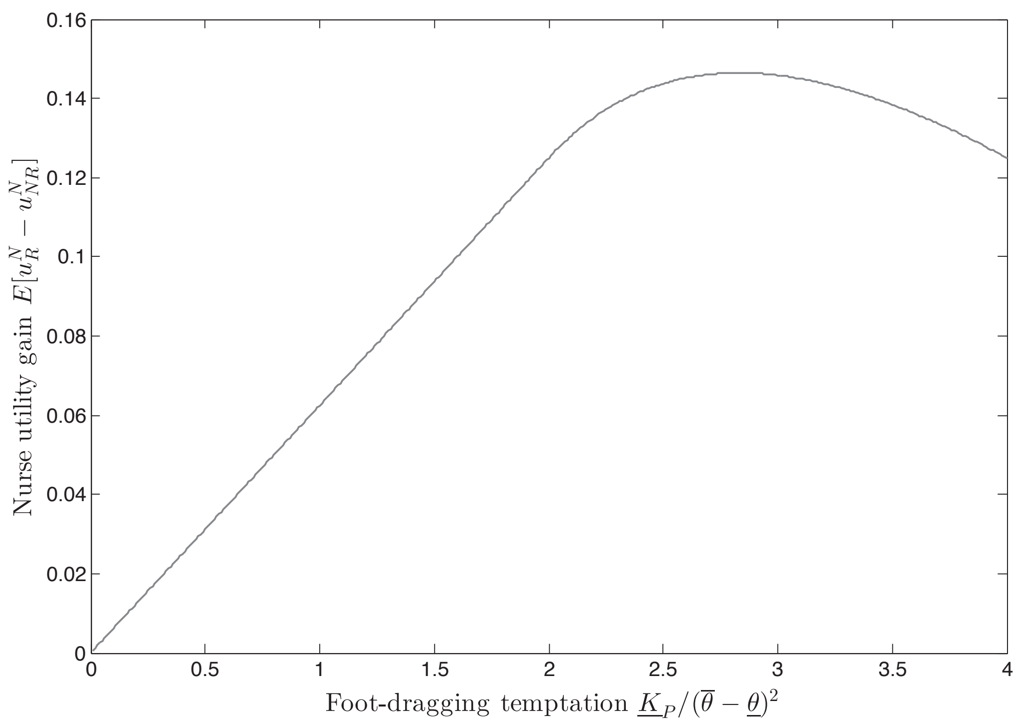
Rubinstein, Ariel, "Perfect Equilibrium in a Bargaining Model," *Econometrica*, 1982, *50* (1), 97–109.

Figure A-2.1: Assignment Policy π Depending on Physician Reporting



Note: This figure shows the assignment policy π, depending on whether physicians can report their types to the triage nurse. Under both cases, I assume that the triage nurse can commit to an assignment policy function. Under no physician reporting, I also consider the triage nurse utility function in which she only cares about patient assignment. On the horizontal axis is a summary statistic for the foot-dragging temptation, $\underline{K}_P/\left(\overline{\theta}-\underline{\theta}\right)^2 \in [0,4]$. Note that when $\underline{K}_P/\left(\overline{\theta}-\underline{\theta}\right)^2 = 4$, a low-type physician should fully foot-drag at $\underline{t}^* = \overline{\theta}$. The dashed line plots the assignment policy when physicians cannot report, $\pi^*_{N|D=0}$ given in (A-2.8). The solid line plots the assignment policy when physicians can report, $\pi^*_R = \pi^*_{IT}$, which equals the assignment policy implied by the intertemporal constraint given in Equation (A-2.12). Note that $\pi^*_{N|D=0} = \pi^*_{IC}$, the assignment policy implied by the standard incentive compatibility constraint with "full" lying, given in Equation (A-2.10); however, the former is a function of censuses, while the latter is a function of reported types.

Figure A-2.2: Efficiency Gain in *ex ante* Assignment with Physician Reporting



Note: This figure shows the gain in *ex ante* assignment efficiency, or the gain in triage nurse utility when she only cares about assignment, that occurs with physician reporting. I compare expected utility for the triage nurse, $\mathbb{E}\left[u^N\right]$, under no physician reporting and under physician reporting. In both cases, her expected utility is simply $\Pr\left\{\theta_{j(3)} = \overline{\theta}\right\}$, normalizing $\overline{K}^N = 1$. Under both cases, I assume that the triage nurse can commit to an assignment policy function. On the horizontal axis is a summary statistic for the foot-dragging temptation, $\underline{K}_P/\left(\overline{\theta}-\underline{\theta}\right)^2 \in [0,4]$. Note that when $\underline{K}_P/\left(\overline{\theta}-\underline{\theta}\right)^2 = 4$, a low-type physician should fully foot-drag at $\underline{t}^* = \overline{\theta}$. On the vertical axis, I plot the difference in *ex ante* assignment efficiency between physician reporting and no reporting.

Appendix B (for Online Publication)

B-1 Pod Trends

In Section 3.2 of the main paper, I introduce two locations, or "pods," within the ED that I use to estimate the overall effect of the self-managed system. Alpha pod always operated as a self-managed system, while Bravo pod switched from a nurse-managed system to a self-manged one in March 2010. I describe in that section that Alpha generally received the more time-intensive and complex patients, because it has always been opened 24 hours, but that Bravo began receiving more complex patients over time, in part due to an increase in the overall volume and complexity of patients arriving at the ED.

In this section, I show evidence of these trends. Understanding these trends is relevant for assessing the robustness of causal inference of the overall effect. If over time Bravo received less time-intensive patients according to unobservable characteristics, then I would not be able to distinguish the effect of the self-managed system from patient selection. On the other hand, if Bravo received more unobservably time-intensive patients over time, then I would instead have a conservative estimate for the causal effect of the self-managed system.

The primary approach I take in this appendix is to examine trends in observable patient and physician peer characteristics, even though I control for observable characteristics in estimating the overall effect in Section 4 of the main paper. If increasingly complex and time-intensive patients are assigned to Bravo pod over time according to observable characteristics, then I will have greater confidence that unobservable characteristics are not going in the opposite direction. This approach complements the approach I take in the main paper of using frequent observations in a long time span to show conditionally parallel trends between the pods and robustness to including pod-specific trends, both of which should include unobservable factors at the pod-time level.

B-1.1 Patient Characteristics

Table B-1.1 presents average patient characteristics for patients in Alpha and in Bravo. It shows that older patients and patients with more severe conditions (a lower emergency severity index indicates a more severe condition) were generally sent to Alpha.

Figures B-1.1 to B-1.3 plot average patient age, Emergency Severity Index (ESI), and number of Elixhauser indices, respectively, for patients seen in Alpha and Bravo over time. The ESI is an integer ranging from 1 to 5, which is the product of a triage algorithm based on patient pain level, mental status, vital signs, and medical condition (Tanabe et al., 2004). An ESI of 1 represents the most severe patient, while an ESI of 5 represents the least severe patient. Elixhauser indices capture 30 important medical conditions, based on coded diagnoses in the medical record, including congestive heart failure, diabetes, hypothyroidism, AIDS, metastatic cancer, and drug abuse (Elixhauser et al., 1998). In regressions in the main paper, I include dummies for each one of these conditions. In Figure B-1.3, I simply plot the number of Elixhauser indices. These three figures are consistent with more complex and severe patients being sent to Alpha at baseline and with increasingly complex and severe patients being sent to Bravo over time.

Figure B-1.4 summarizes the combined effect of observable patient characteristics on patient length of stay by plotting predicted log length of stay. I first estimate the following regression, using only visits to Alpha pod in 2005:

$$Y_{it} = \beta \mathbf{X}_{it} + \varepsilon_{ijkt}, \tag{B-1.1}$$

where $\mathbf{X}_{it}$ are rich patient characteristics that can include the following: age, sex, race, language, ESI, Elixhauser indices, and Major Diagnostic Categories (25 mutually exclusive categories generally based on the organ system and determined by the primary diagnostic code) for patient i at visit arrival time t. I then use estimates $\hat{\beta}$ from Equation (B-1.1) to generate expected lengths of stay for all patients. Results in Figure B-1.4 are predictions based on age, sex, and ESI. Other predictions based on more inclusive sets of patient characteristics, potentially endogenous because they are in part based on physician coding, but they show similar relationships. Patients with longer predicted lengths of stay are always sent to Alpha, but this differential reduces over time.

B-1.2 Physician and Peer Characteristics

In also examine the characteristics of physicians working in Alpha versus Bravo over time. Note that an important feature of the empirical setting is that I observe the same physicians and other providers in both pods over time. This allows me condition on workers in estimating the effect of the self-managed system. Nonetheless, in this section, I explore whether there are any systematic trends in assigning physicians and peers to Alpha versus Bravo over time. This complements Section B-2, which shows that the overall effect is robust to including peer characteristics (in addition to physician-nurse-resident identities), and Section B-6, which shows that physicians are as good as randomly assigned to patient and peer types.

The first set of results are based on physician productivity. Physician productivity is first estimated as fixed effects in a regression of length of stay, controlling for all possible interactions of other team members (physician assistant or resident and nurse), coworker, pod location; patient demographics (age, sex, emergency severity index, Elixhauser comorbidities); ED arrival volume; and time dummies (month-year combination, day of the week, and hour of the day). The average difference in productivity between physicians of above- and below-median productivity is 0.28, meaning that physicians with above-median productivity take 28% less time than those with below-median productivity.

Figure B-1.5 shows average physician productivity for each pod, month, and year combination. Figure B-1.6 shows similar productivity averages for peers, when present. Both figures show an increase in the fixed effect over time for both pods, which is an artifact of the fact that physicians observed during earlier times are observed for a longer time period, which results in lower average lengths of stay. Figure B-1.7 shows averages in the difference in productivity between physicians and peers, when peers are present. Note that a difference in average physician fixed effects of 0.01 between pods implies that physician identities explains a 1% difference in length of stay between the two pods. Given results in B-7.1, a difference in average peer fixed effects of 0.01 would lead to an even lower 0.1% difference in length of stay. There are no economically significant differences in physician productivity trends between pods, compared to the overall effect of the self-managed system on length of stay of -11% to -15%.

The second set of results are based on physician tenure, which is calculated as the difference between the patient date of visit and the physician date of hire. Figure B-1.8 shows average physician tenure for each pod, month, and year combination. Figure B-1.9 shows the corresponding tenure for peers, when present. Figure B-1.10 shows the difference between physician and peer tenure, when a peer is present. There does not appear to be any substantial difference in trends between Alpha and Bravo for any of these tenure measures.

B-1.3 Average Log Length of Stay

I finally plot out average, unadjusted log lengths of stay for both pods and each month in Figure B-1.11. These unadjusted numbers show a gradually increasing trend in length of stay in Bravo and also a decreasing trend in Alpha. These trends occur prior to the Bravo regime change in March 2010. Alpha's trend is continuous across the regime change. Although it is difficult to spot a break in Bravo's trend in unadjusted log length of stay, there does appear to be a decrease in average length of stay starting in February 2010. Recall that the regime change was announced in January 2010.

Results in Section B-1.1 suggest that at least part of this is due to more intensive patients (i.e., patients expected to stay longer) assigned to Bravo over time. While these trends are not unconditionally parallel, I show in Section 4 and Figure 2 in the main paper that they are quite parallel when controlling for patient characteristics and provider identities. Moreover, the differential trends in unadjusted log length of stay goes in the opposite direction of the estimated self-managed effect in the regression, which is that switching to a self-managed system *reduced* Bravo's length of stay.

B-2 Robustness of Overall Effect to Peers

In Section 4 of the main paper, I estimate the overall effect of the self-managed system conditioning on increasingly rich sets of patient characteristics. I find that the effect on length of stay increases in magnitude from -11% to -13% as I include more patient controls and increases further to -15% when I allow for pod-specific time trends. This is consistent with the fact, shown in Section B-1.1, that increasingly more severe and complex patients were sent to Bravo, relative to Alpha, over time.

In contrast, Section B-1.2 shows no qualitatively significant difference in trends in peer characteristics between the two pods. In this section, I formally show that controlling for peer characteristics has no economically significant influence on the estimated effect of the self-managed system. Augmenting Equation (4.1), I estimate

$$\begin{aligned} Y_{ijkpt} &= \alpha Self_{pt} + \sum_{s=1}^{3} 1\left(PeerState_{jt} = s\right)\left(\gamma_s + \left(1 - NoPeer_{jt}\right)\delta \mathbf{PeerChar}_{jt}\right) + \\ &\quad \beta \mathbf{X}_{it} + \eta \mathbf{T}_t + \zeta_p + \nu_{jk} + \varepsilon_{ijkpt}, \end{aligned}$$

where $PeerState_{jt} \in \{1, 2, 3\}$ is represents one of three possible states (peer present in pod, no peer in pod but another physician in ED, and alone in ED), $NoPeer_{jt}$ is a dummy that equals 1 if there is no peer present, and $\mathbf{PeerChar}_{jt}$ is a vector of characteristics for the peer of physician j, which could matter if there is a peer present. This vector includes the cumulative number of days that physician j and his peer have worked together, the tenure of the peer, the difference between the physician's and peer's tenures, the peer's productivity fixed effect, and the difference between the physician's and peer's productivity fixed effects. The coefficient of interest, α, is essentially unchanged at -0.133 with a robust standard error of 0.041, yielding a p-value of 0.002.

B-3 Distribution of Physician Productivity

In this appendix, I briefly discuss three methods of estimating the distribution of physician effects on length of stay, taking into consideration the fact that physician effects are estimated with measurement error. Consistent with regressions in the main text, consider the following regression of log length of stay:

$$Y_{ijkpt} = \beta \mathbf{X}_{it} + \eta \mathbf{T}_t + \zeta_p + c_j + \nu_k + \varepsilon_{ijkpt}, \tag{B-3.1}$$

for patient i, physician j, nurse-resident k, pod p, and patient arrival time t. As before, I control for patient characteristics $\mathbf{X}_{it}$, a vector of time characteristics (hour of the day, day of the week, and month-year interactions) $\mathbf{T}_t$, and pod fixed effects ζ_p. While I previously partialed out physician-nurse-resident trio identities with a term ν_{jk}, I now am separately interested in the physician effects with the term c_j (and now partial out nurse-resident effects with ν_k).

In the standard fixed effect estimation, c_j is estimated with error:

$$\hat{c}_j = c_j + \xi_j,$$

where $\hat{c}_j$ is a measured physician effect, c_j is the true physician effect, and ξ_j is an error term. The fixed effect estimator is an unbiased estimate of c_j, but because of ξ_j, the standard deviation of the distribution of $\hat{c}_j$ will overestimate the standard deviation of the true distribution of c_j, which is the object of interest. All three methods assume that $c_j \sim N\left(0, \sigma_c^2\right)$ and attempt to recover σ_c^2.

B-3.1 Empirical Bayes Estimator

The preferred method uses an empirical Bayes (EB) procedure, based on Morris (1983). This procedures assumes that $\xi_j \sim N\left(0, \pi_j^2\right)$, or equivalently $\hat{c}_j \left| c_j, \sigma_\xi^2 \right. \sim N\left(c_j, \pi_j^2\right)$. For a prior distribution $c_j \sim N\left(\bar{c}, \sigma_c^2\right)$, the posterior distribution that conditions on estimates $\hat{c}_j$ from fixed effect estimation and σ_ξ^2 is

$$c_j \left| \hat{c}_j, \sigma_c^2, \sigma_\xi^2 \right. \sim N\left(c_j^{EB}, \sigma_\xi^2\left(1 - B_j\right)\right),$$

where $c_j^{EB} = \left(1 - B_j\right)\hat{c}_j + B_j\bar{c}$ and $B_j = \pi_j^2 / \left(\pi_j^2 + \sigma_c^2\right)$. B_j "shrinks" $\hat{c}_j$ towards the prior mean of $\bar{c}$, the extent of which depending on the degree of measurement error π_j^2.

As described in Morris (1983), I implement the following feasible version of the procedure:

1. Estimate $\hat{c}_j$ in Equation (B-3.1) by fixed effects estimation. The estimated $\hat{c}_j$ will also have a standard error, which is squared to yield a value called $\hat{\pi}_j^2$.

2. Construct a set of weights for each physician, denoting the weight for physician j as W_j. Begin the procedure with $W_j = 1$ for all j.

3. Iterate the following to convergence:

   (a) Estimate $\bar{c} = \sum_j \left(W_j \hat{c}_j\right) / \sum_j W_j$ and

   $$\hat{\sigma}_c^2 = \frac{\max\left\{0, \sum_j W_j \left[\left(\hat{c}_j - \bar{c}\right)^2 - \hat{\pi}_j^2\right]\right\}}{\sum_j W_j}.$$

   (b) Recalculate $W_j = 1 / \left(\hat{\pi}_j^2 + \hat{\sigma}_c^2\right)$.

I take the final value of $\hat{\sigma}_c^2$ as the EB-adjusted measure of the variance of the distribution of c_j. The standard deviation of this distribution from EB estimation is $\hat{\sigma}_c^{EB} = 0.091$. This implies that increasing true physician productivity by one standard deviation would result in a 9.1% improvement in length of stay.

B-3.2 Random Effects Estimator

The second method is by random effects estimation, in which I assume that c_j is drawn from a random distribution with variance σ_c^2, and I directly estimate $\hat{\sigma}_c^2$, by maximum likelihood estimation. In order to avoid the incidental parameters problem, I first reduce the dimensionality of the covariates by estimating a OLS-predicted length length of stay based on the covariates. The procedure is thus as follows:

1. Estimate $\hat{Y}_{ikpt}$, predicted log length of stay without conditioning on physician j, by the equation $Y_{ijkpt} = \beta \mathbf{X}_{it} + \eta \mathbf{T}_t + \zeta_p + \nu_k + \tilde{\varepsilon}_{ijkpt}$.

2. Estimate by maximum likelihood

$$Y_{ijkpt} = \beta \hat{Y}_{ikpt} + c_j + \varepsilon_{ijkpt}, \tag{B-3.2}$$

   assuming that c_j is randomly distributed as $c_j \sim N\left(0, \sigma_c^2\right)$.

By this procedure, the estimated standard deviation of c_j is $\hat{\sigma}_c^{RE} = 0.065$.

The key random effects assumption is that

$$E\left[c_j \left| \mathbf{X}_{it}, \mathbf{T}_t, \zeta_p, \nu_k \right.\right] = 0, \tag{B-3.3}$$

which means that c_j for physician j is assumed to be orthogonal to patient types, time periods, pods, and nurse-resident teams that the physician is observed to be associated with (e.g., Wooldridge, 2010b). In this case, given the above reduction of dimensionality and the assumption that Y_{ijkpt} takes the form of Equation (B-3.2), the assumption in Equation (B-3.3) takes the following weaker form:

$$E\left[c_j \left| \hat{Y}_{ikpt} \right.\right] = 0. \tag{B-3.4}$$

In Section B-6, I show that physicians are as good as randomly assigned to patient types, peer types, and ED conditions such as patient volume to the ED, *conditional* on time periods and pod locations. It is not generally true that physicians are as good as randomly assigned to work at different times and in different pods. Rather, physicians are allowed to state preferences, which are followed to some degree, even though I observe physicians working in both pods and in all days of the week and hours of the day. Thus, any correlation between physician identities, pods, and time categories would cause the random effects estimation to be biased.

B-3.3 Correlated Random Effects Estimator

The third method allows for correlated random effects (e.g., Mundlak, 1978; Altonji and Matzkin, 2005; Wooldridge, 2010a), which parametrically model the correlation between c_j and covariates. Specifically, I assume $c_j = u_j + v_j = \lambda \hat{E}\left[\hat{Y}_{ikpt}\middle| j\right] + v_j$, where

$$E\left[v_j \left| \hat{Y}_{ikpt} \right.\right] = 0. \tag{B-3.5}$$

This model considers a "fixed" component to the physician effect, u_j, that is predicted by exposure to covariates, namely average predicted log length of stay, and a random component, v_j, that is assumed orthogonal to $\hat{Y}_{ikpt}$.

I implement this estimation by the following:

1. Estimate $\hat{Y}_{ikpt}$, predicted log length of stay without conditioning on physician j, by the equation $Y_{ijkpt} = \beta \mathbf{X}_{it} + \eta \mathbf{T}_t + \zeta_p + \nu_k + \tilde{\varepsilon}_{ijkpt}$.

2. Calculate empirical expectations $\hat{E}\left[\hat{Y}_{ikpt}\middle| j\right] = \sum_{j(i,t)=j}\left(s_{it}\hat{Y}_{ikpt}\right) / \sum_{j(i,t)=j} s_{it}$, where s_{ikpt} is an indicator function for whether $\hat{Y}_{ikpt}$ exists (i.e., all covariates used in step 1 are non-missing) and $j\left(i,t\right)$ is an assignment function indicating the physician associated with patient i arriving at time t.

3. Estimate by maximum likelihood

$$Y_{ijkpt} = \beta \hat{Y}_{ikpt} + \lambda \hat{E}\left[\hat{Y}_{ikpt}\middle| j\right] + v_j + \varepsilon_{ijkpt}, \tag{B-3.6}$$

   assuming that v_j is randomly distributed as $v_j \sim N\left(0, \sigma_c^2\right)$.

The empirical standard deviation of $\hat{u}_j \equiv \hat{\lambda}\hat{E}\left[\hat{Y}_{ikpt}\middle| j\right]$, weighted by $\sum_{j(i,t)=j} s_{it}$, is $\hat{\sigma}_u = 0.038$. I estimate the standard deviation of v_j as $\hat{\sigma}_v = 0.056$. Given that u_j and v_j are orthogonal by assumption, the standard deviation of the physician effects $c_j = u_j + v_j$ is $\hat{\sigma}_c^{CRE} = \sqrt{\hat{\sigma}_u^2 + \hat{\sigma}_v^2} = 0.068$.

B-3.4 Summary of Estimators

In summary, I have estimated three different measures of variation in physician-level effects c_j in Equation (B-3.1). The EB estimator suggests a standard deviation of $\hat{\sigma}_c^{EB} = 0.091$ for the distribution of physician effects. Random effects and correlated random effects estimators suggest $\hat{\sigma}_c^{RE} = 0.065$ and $\hat{\sigma}_c^{CRE} = 0.068$, respectively. The random effects estimator suffers from the problem that c_j may be correlated with other covariates that predict length of stay, while the correlated random effects estimator partially addresses this by allowing for parametric correlation between average covariates and a fixed portion u_j of c_j while allowing for a random v_j that is uncorrelated with covariates. Considering the EB estimator as my preferred (and conservative) measure of variation in physician productivity, an overall effect of 11-15% for the self-managed system is equivalent to 1.2-1.7 standard deviations of physician productivity.

B-4 Alternative Methods of Inference

In estimating the overall effect of the self-managed system in Section 4 of the main paper, my baseline specification clusters standard errors by physician, which is equivalent to an experiment sampling at the level of physicians, who are given shifts that translate to pods and organizational systems, before and after the regime change in Bravo. This thought experiment is supported by evidence of conditional quasi-random assignment of physicians to patients and peers, shown in Section B-6.

In Sections B-4.1 and B-4.2, I discuss two alternative methods of inference that allow for pod-level random shocks, given that I have only two pods and cannot cluster by pod for inference. The intuition underlying both of these alternative methods takes advantage of the fact that I observe a long time dimension for both pods. I can therefore compare "random" variation in the difference between the pods over many points in time, which are not associated with any regime change, with the effect associated with Bravo pod's switch from a nurse-managed to self-managed system in March 2010.

B-4.1 Inference with Serially Correlated Pod-level Error Terms

In this approach, I address sampling variation at the pod level across time with a parametric form on the error terms. Specifically, I allow for pod-month random shocks that are serially correlated across months by a first-order autoregressive (AR1) process. That is, I first calculate month-year-pod fixed effects from

$$Y_{ijkpt} = \sum_{m=1}^{M} \sum_{y=1}^{Y} \alpha_{myp} I_{t \in m} I_{t \in y} + \beta \mathbf{X}_{it} + \tilde{\eta} \tilde{\mathbf{T}}_t + \nu_{jk} + \varepsilon_{ijkpt}, \tag{B-4.1}$$

which is the same model used in the paper to generate Figure 2. $I_{t\in m}$ and $I_{t\in y}$ are indicator functions for t belonging in month m and year y, respectively, and $\tilde{\mathbf{T}}_t$ is a revised vector of time categories only for day of the week and hour of the day. Coefficients estimated for α_{myp} are used as data points.

In the second stage, I estimate a model with observations collapsed to the month-year-pod level:

$$\hat{\alpha}_{myp} = \gamma \mathit{Self}_{myp} + \eta_{my} + \zeta_p + \varepsilon_{myp},$$

where $\hat{\alpha}_{myp}$ are estimated coefficients from Equation (B-4.1), γ is the coefficient of interest on the treatment indicator Self_{myp} of whether pod p is a self-managed system during month m and year y, η_{my} and ζ_p are respective time and pod fixed effects, and ε_{myp} is a serially correlated error term with the AR1 process

$$\varepsilon_{myp} = \rho \varepsilon_{my-1,p} + z_{myp}.$$

I estimate a self-managed effect of $\hat{\gamma} = -0.121$, with a standard error of 0.025, which is significant with a p-value less than 0.001. This estimate is quite similar to the baseline estimate in the main paper. The estimated correlation in error terms across months is $\hat{\rho} = 0.184$.

B-4.2 Inference with Systematic Placebo Tests (Randomization Inference)

Another alternative method of inference takes sampling as fixed and instead considers randomization at the level of treatment (Rosenbaum, 2002). This is also in the same spirit of systematic placebo tests (e.g., Abadie et al., 2010). Under the sharp null of no effect of the self-managed system, there should be no significant difference between my obtained estimates and those I would obtain if I consider a number of placebo regime changes over each pod and month.

Here, I again use estimated month-year-pod fixed effects $\hat{\alpha}_{myp}$ from Equation (B-4.1). I then perform a regression at placebo regime changes at each month and pod with a bandwidth of six months prior to the placebo regime change date and six months post. That is, using only only observations from within the bandwidth for the two pods, I estimate a difference-in-differences regression

$$\hat{\alpha}_{myp} = \gamma^r \mathit{PlaceboSelf}^r_{myp} + \eta_{my} + \zeta_p + \varepsilon_{myp},$$

where $\mathit{PlaceboSelf}^r_{myp}$ is an indicator for whether pod p at month m and year y is self-managed under the placebo regime change r.

Out of the 120 estimates for γ^r, the true regime change in March 2010 had the second-largest coefficient of $\gamma^{\text{Mar2010}} = -0.131$, corresponding to a randomization-inference p-value of about 0.008. Notably, the second largest coefficient corresponds to February 2010, with $\gamma^{\text{Feb2010}} = -0.128$, the month before the official regime change. The regime change was announced in January 2010; as shown in Section B-8, there appears to be some anticipatory response by physicians in Bravo beginning to see more patients in beds outside of their own (e.g., physicians in Bravo-1 shifts seeing patients in beds belonging to Bravo-2). This result, including the ranking of coefficients, is robust to using different bandwidths (significance is slightly greater with lower bandwidths).

B-5 Overall and Pod Patient Volume

Patient volume, or the flow of patients, is a key feature of work in the ED. Although differences in patient volume are expected across different hours of day at the time of scheduling, there remains substantial additional variation in realized patient volume within time categories. Furthermore, while working in a pod, although physicians are aware via the computer interface of the current state of patient volume, there remains substantial uncertainty about the distribution of these patients to pods. I use both of these facts in order to identify the mechanism of foot-dragging as the response to expected future work.

B-5.1 Time Category Variation, Across and Within

I first calculate overall patient volume as the number of patients arriving to the waiting room at each hour in the data sample, including hours with no patients arriving. I then calculate summary statistics for patient volume – extrema and quantiles – across observations within each value of a time category. Figures B-5.1 to B-5.3 show respective variation relevant to hours of the day, days of the week, and month-year interactions. These figures are meant to represent qualitative variation across and within time categories. In essentially all regressions in the paper, I include fixed effects for each value of each time category. In Section B-6.1, I also show evidence of quasi-random exposure to patient volume, conditional on the time categories.

As shown in Figure B-5.1, perhaps the greatest variation across time categories is attributable to hours of the day. As expected, reflected in staffing schedules, more patients arrive during the day than at night. However, there remains significant variation within each hour of the day, with the 95th percentile always more than double the volume of the 5th percentile even at peak hours. Figure B-5.2 shows variation across and within days of the week. Monday is the busiest day, but the variation across days of the week is negligible compared to the variation that remains after accounting for day of the week. Similarly, as shown in Figure B-5.3, variation across month-year observations is negligible compared to within month-year variation. There does appear to be a trend towards greater patient volumes, especially in above-median quantiles, over time.

B-5.2 Low Correlation between Overall and Pod Volume

Figure B-5.4 presents plots that show the relationship between the number of patients arriving at triage in the hour prior to the index patient's pod arrival, which I use in the main manuscript as a measure of expected future work. In order to show that I can separate expected future work from actual (current or future) work, I show that there is wide variation in pod-specific patient volume, even conditional on overall ED volume arriving at triage. Each of the four panels in Figure B-5.4 examines a different hour for pod-specific volume. The median time from triage to pod is about 30 minutes, and the median time from pod to discharge order is about 3.4 hours.

Correlation coefficients are low: The greatest correlation coefficient coefficient is 0.21, between overall volume and pod-specific volume in the same hour previous to the index patient's arrival to the pod. Correlation coefficients between previous-hour overall volume and pod-specific volume for the subsequent hour, the hour ending two hours later, and the hour ending three hours later are 0.16, 0.13, and 0.12, respectively.

B-6 Conditional Quasi-random Exposure to Patients and Peers

Although the effect of the self-managed system is identified by the Bravo regime change, providers working in both pods over time, and conditionally parallel trends in the two pods, it may also be useful to show evidence of conditional quasi-random exposure of physicians to patients and peers. First, this addresses more complicated threats to identification that involve both physician selection and time-varying productivity. Second, it is consistent with the sampling thought experiment in which physicians are assigned random pod experiences conditional on rough time categories. Third, quasi-random exposure to patient volume in particular supports the idea of exogenous shocks to expected future work, which I use to identify foot-dragging.

B-6.1 Similar Exposures across Physician Types

I first investigate whether physicians of different types are exposed to similar patient types and ED volume conditions. I focus on physician differences in terms of preferences and productivity. I estimate preferences for specific patient types by the probability that a physician will choose a patient type when given the choice. I form productivity measures by fixed effects for physician identities in a regression of log length of stay. Physicians with one standard deviation greater preferences for a patient type are 7.4% more likely to choose that patient type than average. Physicians who are one standard deviation faster than average have 11% shorter lengths of stay.

In Table 1 in the main paper and Table B-6.1, I show that physicians that differ by productivity or preferences, respectively, are exposed to similar average patient types and ED patient volume. In addition to showing that averages are similar, I examine the distribution of patient volume more closely. Figure B-6.1 shows that the distribution of patient volume for high- and low-productivity physicians are indistinguishable.

B-6.2 Joint Insignificance of Physician Identities

I test for the joint significance of physician identities in regressions of patient characteristics arriving at the pod and ED patient volume, while conditioning on rough indicators of time. For patient characteristics, I first summarize patient characteristics of age, sex, emergency severity index (ESI), race, and language into predicted length of stay. I calculate this prediction for each patient and then average these predictions for patients arriving at each pod and hour-date in my data. Using physician schedules, I associate the average predicted length of stay for each pod-hour combination to physicians that are working on that pod and during that hour-date. I then estimate this equation:

$$\overline{E\left[Y\right]}_{pt} = \alpha_j I_{jpt} + \eta \mathbf{D}_{pt} + \varepsilon_{jpt},$$

where $\overline{E\left[Y\right]}_{pt}$ is the average predicted log length of stay for pod p at time (hour-date combination) t, I_{jpt} is an indicator variable equal to 1 if physician j is working in pod p at time t, and $\mathbf{D}_{pt}$ is a vector of interactions between pod p and rough time dummies including month-year dummies, day of the week dummies, and hour of the day dummies. I test for the joint significance of the vector of coefficients $\boldsymbol{\alpha} = (\alpha_j)$. The F-statistic (64, 659042) under the null that $\boldsymbol{\alpha} = \mathbf{0}$ is 1.07 (p-value of 0.32), clustering by date.

For ED volume, I perform a similar exercise. I estimate this equation:

$$EDWork_t = \alpha_j I_{jt} + \eta \mathbf{T}_t + \varepsilon_{jt},$$

where $EDWork_t$ is the number of patients arriving at the ED during time (hour-date combination) t, I_{jt} is an indicator variable equal to 1 if physician j is working in the ED at time t, and $\mathbf{T}_t$ is a vector of indicators for rough time categories of month-year, day of the week, and hour of the day. The F-statistic (64, 2098) under the null that $\boldsymbol{\alpha} = \mathbf{0}$ is 1.11 (p-value of 0.26) when clustering by unique date (each of the 75 physicians in the sample works from 60 to 845 days).

B-6.3 Exogenous Assignment of Physicians to Peers

I also show that physician identities do not explain the preferences or ability of their peers that they happen to be working with on a shift. With respect to physician productivity, I regress the productivity (length of stay) fixed effect of the physician against that of his peer. That is, if $\hat{\alpha}_j$ is the productivity fixed effect for physician j, and $-j\,(t)$ denotes physician j's peer during shift t, I perform this regression:

$$\hat{\alpha}_{-j(t)} = \beta\hat{\alpha}_j + \varepsilon_{jt}.$$

I estimate the coefficient β to be small and insignificant at -0.003 with a standard error of 0.014 (p-value of 0.84). Similarly, the correlation coefficient between physician and peer fixed effects is -0.018 (p-value 0.16).

If I normalize the average productivity fixed effect of peers working with physicians who are faster than average to be 0 (with standard deviation 0.107), the average productivity fixed effect of peers working with physicians who are slower than average is 0.0001 with standard deviation of 0.105.

If instead of entering the physician's productivity fixed effect on the right-hand-side, I enter physician dummies in the specification

$$\hat{\alpha}_{-j(t)} = \eta_j + \varepsilon_{jt},$$

where the error term is clustered by unique pairings of physician and peer, I am unable to reject the null that the vector $\boldsymbol{\eta}$ is jointly 0 with an F-statistic (53, 1995) of 0.57 (p-value 0.99).

I perform similar analyses with respect to physician preferences and find no relationship between the preferences of peers.

B-7 Peer Effects

This section extends the analysis of peer effects in two ways. First, I estimate direct effects in Section B-7.1, which is the effect of working with a peer of higher productivity. Second, I consider how foot-dragging might depend on the relationship between the peer and the index physician in Section B-7.2.

B-7.1 Direct Peer Effects

While the main paper focuses on peer effects on foot-dragging, as an interaction between the presence of a peer and expected future work, I consider the direct effect of peer productivity in this appendix. Consistent with Mas and Moretti (2009), I find that working with a faster peer shortens lengths of stay for the index physician, but I do not find a significant difference in this peer effect between organizational systems.

The effect of peer productivity may occur through a variety of mechanisms. Social incentives have been discussed as a source of positive peer effects (e.g., Mas and Moretti, 2009). Knowledge spillovers are yet another mechanism for peer effects. In addition, in this setting, strategic behavior may also lead to positive peer effects. For example, in the nurse-managed system, a physician working with a less-productive peer will be more likely to get new work unless if he slows down. Moreover, other mechanisms discussed in the main paper could influence the sign and magnitude of peer effects. Free riding by waiting for productive peers to choose work would have a negative influence on peer effects, while dynamic smoothing and matching may have a positive influence if productive peers have complementary skills and availability or a negative influence if less-productive peers have the complementary skills and availability.

Furthermore, social incentives may differ between the self-managed and nurse-managed systems because the rules of the games differ. In the nurse-managed system, peers impose a negative externality by being less productive (i.e., the foot dragging externality), and social incentives may increase efficiency. In contrast, in the self-managed system, peers actually may impose a positive externality by being less productive to others within the team, because they prevent work from being sent to the pod. In addition to the direction of the social incentives, their strength may differ between the two organizational systems as peers work more closely in self-managed teams.

Although peer effects may be difficult to interpret in terms of mechanisms for these reasons, it is still interesting to compare between the two organizational systems in reduced form. It is certainly possible that some physician-peer combinations may perform better in a self-managed setting while others may perform better in a nurse-managed setting. Employing similar methodology as Mas and Moretti (2009), I first estimate physician fixed effects for log length of stay, and then I use the fixed effects of peers as an explanatory variable in a regression of productivity in order to estimate peer effects.

In the first stage, I estimate the following regression on log length of stay Y_{ijkpt} for patient i, physician j, resident-nurse team k, and visit arrival time t:

$$Y_{ijkpt} = \theta_j + \mathbf{M}\phi_{C_j} + \beta \mathbf{X}_{it} + \eta \mathbf{T}_t + \zeta_p + \varepsilon_{ijkpt}, \tag{B-7.1}$$

where in addition to patient characteristics $\mathbf{X}_{it}$ and time categories $\mathbf{T}_t$, I control for all possible sets of physicians j, resident-nurse combinations k, physician peers l, and pod locations p in order to address the reflection problem (Manski, 1993). These providers and peers are included in a set of all dummies representing possible combinations $\phi_{C_j} = \{C(j,k,l,p)\}$, where

$$C(j,k,l,p) = \begin{cases} 1 & \text{if physician } j \text{is working with team } k \text{and peer } l \text{in pod } p, \\ 0 & \text{if } i = l, \\ 0 & \text{otherwise.} \end{cases}$$

The parameter of interest is the physician fixed effect θ_j. The standard deviation of estimated fixed effects is 0.11, meaning that physicians one standard deviation above mean productivity have lengths of stay that are 11% shorter than average.

In the second stage, I use the set of fixed effects from (B-7.1) in order to estimate the effect of a peer's productivity on the index physician's outcomes, in particular length of stay. Using the fixed effect θ_{-j} for physician j's peer, I estimate this regression using separate samples for the nurse-managed and self-managed systems:

$$Y_{ijkpt} = \alpha\theta_{-j} + \beta\mathbf{X}_{it} + \eta\mathbf{T}_t + \zeta_p + \nu_{jk} + \varepsilon_{ijkpt}. \tag{B-7.2}$$

The coefficient α represents peer effects from working with a peer with productivity θ_{-j}. A positive α suggests that physicians work faster when working with a more-productive (faster) peer and slower when working with a less-productive (slower) peer. I also estimate a pooled version with

$$Y_{ijkpt} = \alpha_1\theta_{-j} \cdot \mathit{Self}_{it} + \alpha_2\theta_{-j} + \alpha_3\mathit{Self}_{it} + \beta\mathbf{X}_{it} + \eta\mathbf{T}_t + \zeta_p + \nu_{jk} + \varepsilon_{ijkpt}.$$

A positive α_2 suggests positive peer effects in the nurse-managed system; a positive α_1 suggests greater peer effects in the self-managed system (in the case of positive peer effects). All regressions require that there be a peer present in the pod.

Table B-7.1 reports estimates of peer effects. A 1% increase in peer productivity leads to a 0.1% increase in physician productivity. Faster peers have a stronger influence than slower peers in both settings. However, peer effect estimates can be relatively imprecise, compared to foot-dragging results in the main paper and in particular in the self-managed system. I cannot reject that overall peer effects are different between the two organizational systems.

While effects are quantitatively similar, they are less precisely estimated in the self-managed system than in the nurse-managed system, despite similar numbers of observations. This suggests that peer effects and perhaps the interaction between physicians in self-managed teams, through a number of possible mechanisms, are generally less predictable. Although I do not show results here, I do not find any significant effect of peer effects of working with a peer with productivity specific to the index patient. That is, working with a physician who is better at seeing heart patients does not improve the productivity of the index physician seeing the heart patient. In addition, I do not find peer effects on quality outcomes, which suggests that the main (and modest) effect of peers is mostly on length of stay. These facts discount the possibility of learning or being helped by more skilled physicians.

B-7.2 Peer Effects on Foot-dragging by Peer Type

Given that foot-dragging is reduced by the presence of a peer, I also consider different types of social relationships between physicians and their peers. I first consider peers of the same sex, similar age, or same place of residency training as potentially more connected to each other. Second, I consider peers who are faster (or more productive) than median. The effect of this peer type on foot-dragging may include both social and strategic concerns. To see the strategic concerns, note that slower peers will cause more work to be redirected to physicians unless they slow down as well. Third, I consider peers, by their history of time working with each other, who are more familiar with each other's workplace behavior and more likely to have established reputations with each other.[1] Finally, I consider peers who have at least two years greater tenure than the index physician. Social hierarchy is a common feature in many workplaces, particularly those with professionals, long tenures, or strong work cultures.

For each of these peer types, I estimate regressions of the following form:

$$\begin{aligned} Y_{ijkpt} &= \alpha_1 EDWork_t + \alpha_2 PeerType_{jt}^m \cdot EDWork_t + \alpha_3 PeerType_{jt}^m + \\ & \quad \beta \mathbf{X}_{it} + \eta \mathbf{T}_t + \zeta_p + \nu_{jk} + \varepsilon_{ijkpt}, \end{aligned} \tag{B-7.3}$$

separately for nurse-managed and self-managed samples. $PeerType_{jt}^m$ is an indicator that the peer for physician j at time t is of type m. I am interested in the coefficient α_2 as the effect of working with a peer type on foot-dragging, again identified by increases in length of stay with respect to expected future work.

Table B-7.2 reports results for three of the peer types.[2] Senior peers are the only peer type showing a significant effect on foot-dragging. In the nurse-managed system, working with a senior peer decreases foot-dragging by half, from an increase of 0.8% for each patient arriving at the ED to an increase of 0.4%. In a pooled regression shown in the third column, it also appears that senior peers further reduce foot-dragging in the self-managed system.[3] Other peer types – highly productive peers, familiar peers, and connected peers – show no significantly differential effect on foot-dragging. These results suggest that the most important social relationships between peers

[1] I use a threshold of at least 60 hours working in the same pod, which is at the 75th percentile, to describe peers that are "familiar" with each other. Given physician turnover and a large number of shift times and locations, it is is relatively uncommon for two ED physicians to have longer histories working together in the same pod.

[2] For brevity, I omit peer types related to social connectedness from Table B-7.2, as they show no effect on foot-dragging.

[3] The pooled regression includes interactions with the self-managed system and a direct effect for the self-managed system. I do not write out this equation above, as Equation (B-7.3) communicates the effect of interest in α_2. Recall that self-management and social incentives may independently reduce foot-dragging, and that there may be some foot-dragging in the nurse-managed system, since the 0 benchmark for no foot-dragging is conservative.

may be unilateral ones based on hierarchy, as opposed to ones that are based on connectedness or familiarity.

B-8 Bed Location of Assigned Patients

The assignment of patients to beds and to physicians is a crucial aspect of the management of work in the ED. As illustrated in Figure 1 and described in the paper, the fundamental difference between the nurse-managed system and the self-managed system arises from the fact that, in the nurse-managed system, physicians "own" beds and are therefore assigned patients who are assigned to their beds, while physicians share beds in the self-managed system. In this section, I describe in greater detail the use of beds, with particular emphasis on Bravo pod, which switched from a nurse-managed system to a self-managed system in March 2010.

B-8.1 Pod Layout and Bed Use

Figures B-8.1 and B-8.2 show a computer interfaces for Alpha and Bravo pods, respectively. Alpha and Bravo had largely similar physical layouts and stable bed locations over time. However, unlike Alpha, Bravo was divided into two administrative zones when it was a nurse-managed system. Two physicians working in Bravo would each be assigned to a zone, Bravo-1 or Bravo-2, which included a set of beds. The zones were physically contiguous and non-overlapping.

While the physical layout of both pods was stable, there may be slight changes in the electronic designations of beds, listed in Table B-8.1. In addition to physical beds, electronically designated beds include virtual beds, such as hallways and other unwalled areas where stretchers can be rolled into.[4] Tables B-8.2 to B-8.4 show the number of patient visits, as well as the dates spanning the first and last observed visit, associated with each bed in Alpha, Bravo-1, and Bravo-2, respectively. Patients are preferentially assigned to physical beds within walled rooms, since these beds have greater privacy, compared to hallway beds. Finally, although the numbers of possible beds in each Bravo zone were roughly equal, Bravo-1 had fewer beds than Bravo-2 that were routinely filled, in part because Bravo-1 had more virtual beds.[5]

B-8.2 Patient Assignment to Beds and to Physicians

The triage nurse assigns all patients from the waiting room to their initial beds within the ED. Once patients are assigned a bed, they appear on the pod-specific computer interface for either

[4] Hallway beds are designated with an "H" and sometimes a number signifying the closest numbered room. Beds 28 to 31 in Bravo-1 are located in unwalled spaces separated by curtains.

[5] Relatedly, Bravo-1 was located closer to the Bravo doors and was reserved to hold patients who would likely exit the ED sooner. In practice, Bravo-1 had patients with shorter unadjusted lengths of stay. Patients were not necessarily less severe in Bravo-1; there was a wider distribution of Emergency Severity Index scores in Bravo-1 versus Bravo-2.

Alpha (Figure B-8.1) or Bravo (Figure B-8.2).[6] Upon being assigned a bed, patients are either technically assigned to physicians, if physicians own beds as in the nurse-managed system, or otherwise must wait for a physician to sign up for their care, when physicians share beds in the self-managed system. In either case, physicians must click on each patient's representation on the computer interface in order to acknowledge that they are assuming the patient's care.

Physicians in the nurse-managed system may not always attend to patients assigned in their zones. Because all patients must be voluntarily acknowledged, physicians working together in the nurse-managed system may still agree to see patients who are not strictly in their zone. This is particularly the case for patients waiting to be seen in virtual beds in the hallway when the pod is busy and for patients near the border between zones, even though all beds (virtual and physical) belong to only one zone. In this sense, the nurse-managed system is never purely managed by the triage nurse, and a small proportion of patients may be assigned by physician discretion.

In addition, patients routinely may change beds during their ED stay. Although the median visit involves only one bed, 44% of patients switch beds at least once. The mean number of beds occupied during a patient visit is 1.85 beds. The 95th percentile for the number of beds occupied is four. Although the triage nurse has sole discretion for the assignment of the initial bed, physicians and nurses may have some input regarding changing beds after initial pod arrival. Physicians are rarely reassigned when patients change beds; in fact, physicians are expected to care for patients during their entire stay and routinely remain present for two to three hours past their end of shift, unless if a patient is expected to stay significantly longer (e.g., greater than three hours) past the end of the responsible physician's shift.

B-9 Additional Results

In this appendix, I present the following additional empirical results, as well as a brief discussion of some of these results:

- Figure B-9.1: Presents the foot-dragging coefficient on expected future work interacted with pod identities and four-month intervals.
- Figure B-9.2: Presents the correlation between censuses and new patient assignment in both pods over time using a local linear regression (Figure 5 in the main text), with additional confidence intervals on estimates.
- Figure B-9.3: Presents the correlation between censuses and new patient assignment in both pods over time using a local linear regression, but extends the use in Equation (7.1)

[6]Patients also appear on the computer interface while they are still in the waiting room. Physicians can tab through each location in the ED, including pods and the waiting room, to observe the patients in each location, their brief clinical information, and the health care providers (if any) assigned to them.

of vestigial "Bravo-1" and "Bravo-2" shift labels until June 2010. Confidence intervals are also shown.

- Figure B-9.4: Presents the correlation between censuses and new patient assignment in both pods over time, but estimates Equation (7.1) with a kernel regression. Confidence intervals are also shown.
- Table B-9.1: Presents the overall effect of the self-managed system on other outcomes, estimated by Equation (4.2), for each of the specifications in Table 2. This table extends Table 3.
- Table B-9.2: Presents responses to expected future work, estimated by Equation (5.2), for outcomes other than length of stay.
- Table B-9.3: Presents responses to expected future work across different peer states, estimated by Equation (6.2) and similar to Table 5, except that expected future work *per physician* is considered.

Figure B-9.1 estimates foot-dragging coefficients, which indicate the response of length of stay to expected future work, interacted with pod identities and four-month intervals. Specifically, I estimate

$$Y_{ijkpt} = \sum_{\tau \in T} \alpha_{p\tau} I_{t\in\tau} EDWork_t + \beta \mathbf{X}_{it} + \eta \mathbf{T}_t + \zeta_p + \nu_{jk} + \varepsilon_{ijkpt},$$

where $\tau \in T$ is a four-month interval. Coefficients of interest, $\{\alpha_{p\tau}\}_{p\in\{0,1\},\tau\in T}$, are specific to pod and four-month interval, but are of course foot-dragging coefficients on expected future work, $EDWork_t$, as measured by the number of patients arriving at the ED in the hour prior to the index patient's bed arrival. I therefore can control for the same rich set of time categories (hour of the day, day of the week, and month-year-interaction) and pod identities. Foot-dragging (i.e., increases in length of stay as expected future work increases) does not immediately disappear in Bravo when Bravo changes from a nurse-managed system to a self-managed system. Foot-dragging is indistinguishable from 0 during all periods in Alpha.

Figure B-9.2 presents the same results as in Figure 5 in the main text, plotting the coefficient on censuses a linear probability model in Equation (7.1) of new patient assignment. The only difference between these two Appendix Figures and Figure 5 is that confidence intervals are plotted for Bravo and Alpha pods, respectively. For most months, I can reject that the central estimate for one pod is the same as the central estimate for the other pod, but the two confidence intervals otherwise overlap. Also, all central estimates for Alpha pod are below those for Bravo pod in the pre-regime period.

Figure B-9.3 plots coefficients and confidence intervals for Bravo pod estimated from the following regression:

$$Y_{ijt} = \alpha Census_{jt} + \beta \mathbf{ShiftTime}_{jt} + \gamma ZoneLabel_{jt} + \eta_j + \nu_{it} + \varepsilon_{ijt}, \tag{B-9.1}$$

which differs from Equation (7.1) only in that it allows for "Bravo-1" and "Bravo-2" shift labels to matter even after the regime change. The reason for this change relative to Equation (7.1) is to consider the possibility that physicians with higher censuses are more likely to be assigned new patients because they are following the norm that they are responsible for beds in Bravo-2, which is the larger zone. Figure B-9.3 is largely the same as Figure 5 in the main text, except that the positive coefficients immediately at the regime change are now reduced. However, there is still a positive coefficient at April 2010 that is close to being significant at the 5% level.

Figure B-9.4 shows estimates of Equation (7.1) using a kernel regression. I use triangular kernels with 45-days on each side, which may be truncated if sufficiently close to March 1, 2010. Results are largely similar to the baseline results, shown in Figure 5 in the main text, which are estimated using a local linear regression. I can reject that the central estimate for one pod is the same as the central estimate for the other pod, but the two confidence intervals otherwise overlap. Also, all central estimates for Alpha pod are below those for Bravo pod in the pre-regime period.

Table B-9.1 shows that results in Table 3 are generally robust to different controls in Equation (4.2). A couple notes for this table are in order. First, the effect on 14-day return visits is generally consistent – a reduction of about 1.2 percentage points from a base of 6.7 percentage points – across specifications, although all of these estimates are only significant at the 10% level. This effect is in the direction of an *improvement* in quality of care, despite reductions in length of stay. Second, log time to first order decreases in Model 5 (controlling for pod-specific time trends). However, this is clinically insignificant, given that average time to first order is only 20 minutes. Recall that I examine time to first order as a proxy for "free-riding" in which physicians delay choosing patients in the self-managed system; a reduction in time to order is inconsistent with free-riding.

References

Abadie, Alberto, Alexis Diamond, and Jens Hainmueller, "Synthetic control methods for comparative case studies: Estimating the effect of California's tobacco control program," *Journal of the American Statistical Association*, 2010, *105* (490), 493–505.

Altonji, Joseph G. and Rosa L. Matzkin, "Cross Section and Panel Data Estimators for Nonseparable Models with Endogenous Regressors," *Econometrica*, July 2005, *73* (4), 1053–1102.

Elixhauser, Anne, Claudia Steiner, D. Robert Harris, and Rosanna M. Coffey, "Comorbidity Measures for Use with Administrative Data," *Medical Care*, January 1998, *36* (1), 8–27.

Manski, Charles F., "Identification of Endogenous Social Effects: The Reflection Problem," *The Review of Economic Studies*, 1993, *60* (3), 531–542.

Mas, Alexandre and Enrico Moretti, "Peers at Work," *The American Economic Review*, 2009, *99* (1), 112–145.

Morris, Carl N., "Parametric Empirical Bayes Inference: Theory and Applications," *Journal of the American Statistical Association*, March 1983, *78* (381), 47–55.

Mundlak, Yair, "On the Pooling of Time Series and Cross Section Data," *Econometrica*, January 1978, *46* (1), 69.

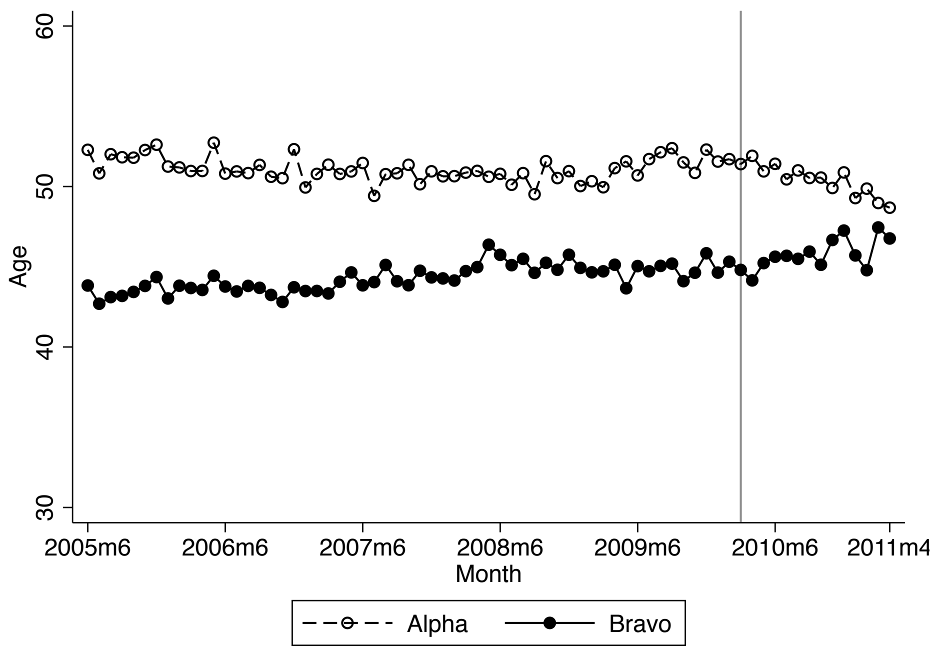
Rosenbaum, Paul, "Covariance adjustment in randomized experiments and observational studies," *Statistical Science*, August 2002, *17* (3), 286–304.

Tanabe, Paula, Rick Gimbel, Paul R. Yarnold, Demetrios N. Kyriacou, and James G. Adams, "Reliability and Validity of Scores on the Emergency Severity Index Version 3," *Academic Emergency Medicine*, 2004, *11* (1), 59–65.

Wooldridge, Jeffrey M., "Correlated Random Effects Models with Unbalanced Panels," Technical Report, Michigan State University 2010.

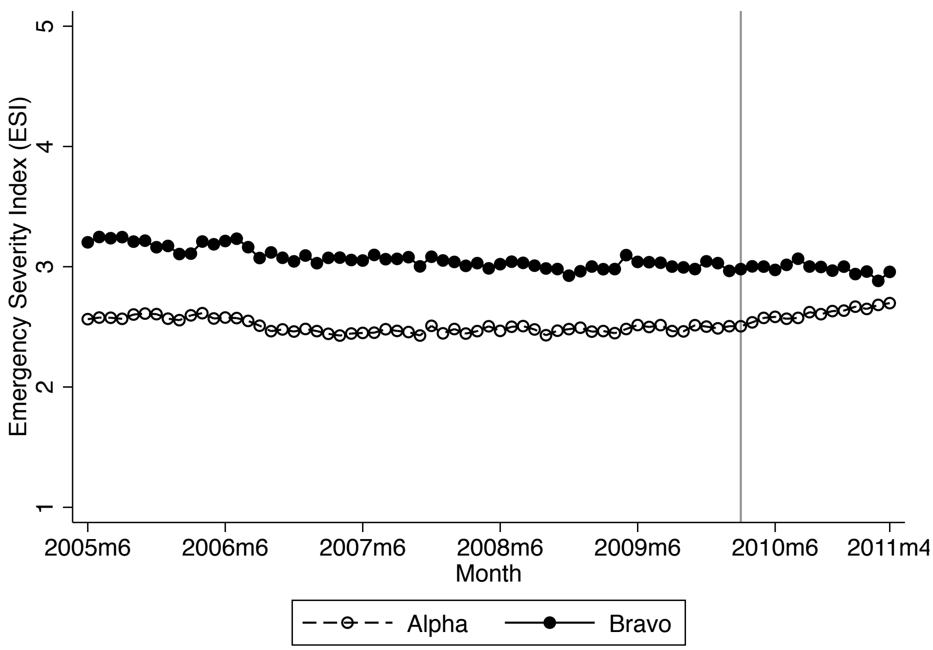
— , *Econometric Analysis of Cross Section and Panel Data*, MIT Press, October 2010.

Figure B-1.1: Average Patient Age



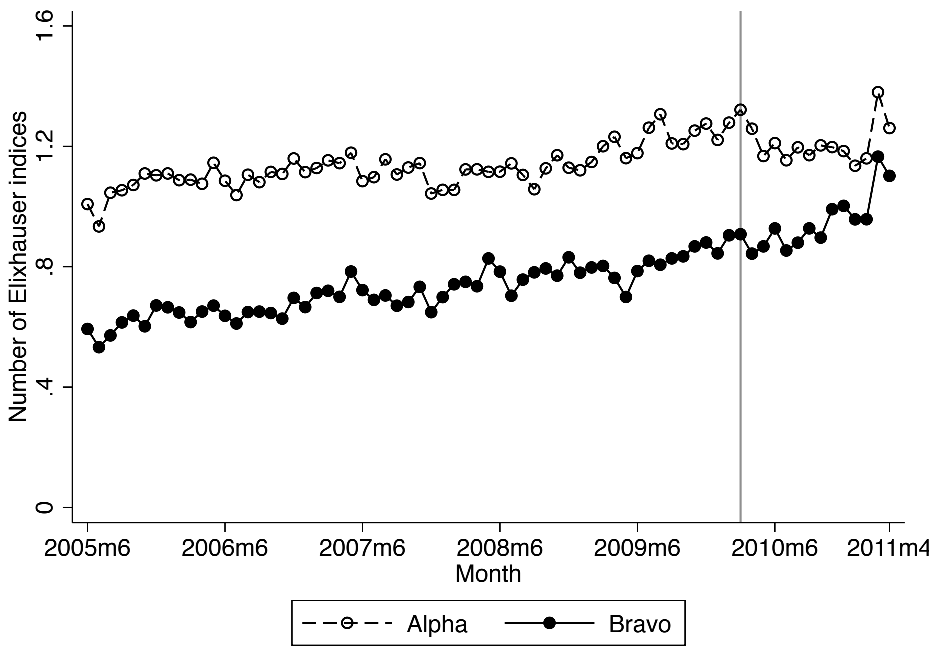
Note: This figure shows average age for patients in each pod, month, and year. Alpha pod averages are plotted with hollow circles; Bravo pod averages are plotted with solid circles. The vertical gray line indicates the month of the regime change of Bravo from a nurse-managed system to a self-managed system, in March 2010. Alpha was always self-managed.

Figure B-1.2: Average Emergency Severity Index



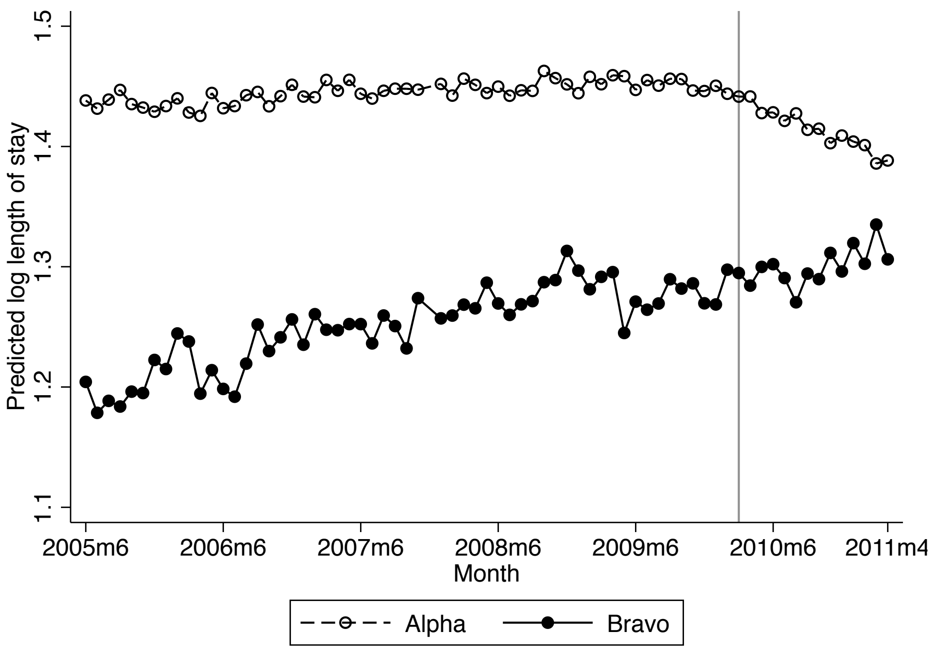
Note: This figure shows average Emergency Severity Index (ESI) for patients in each pod, month, and year. The ESI is a integer from 1 to 5, based on an algorithm accounting for patient pain, mental status, vital signs, and medical condition (Tanabe et al., 2004). An ESI of 1 is the most severe category, while an ESI of 5 is the least severe category. Alpha pod averages are plotted with hollow circles; Bravo pod averages are plotted with solid circles. The vertical gray line indicates the month of the regime change of Bravo from a nurse-managed system to a self-managed system, in March 2010. Alpha was always self-managed.

Figure B-1.3: Average Number of Elixhauser Indices



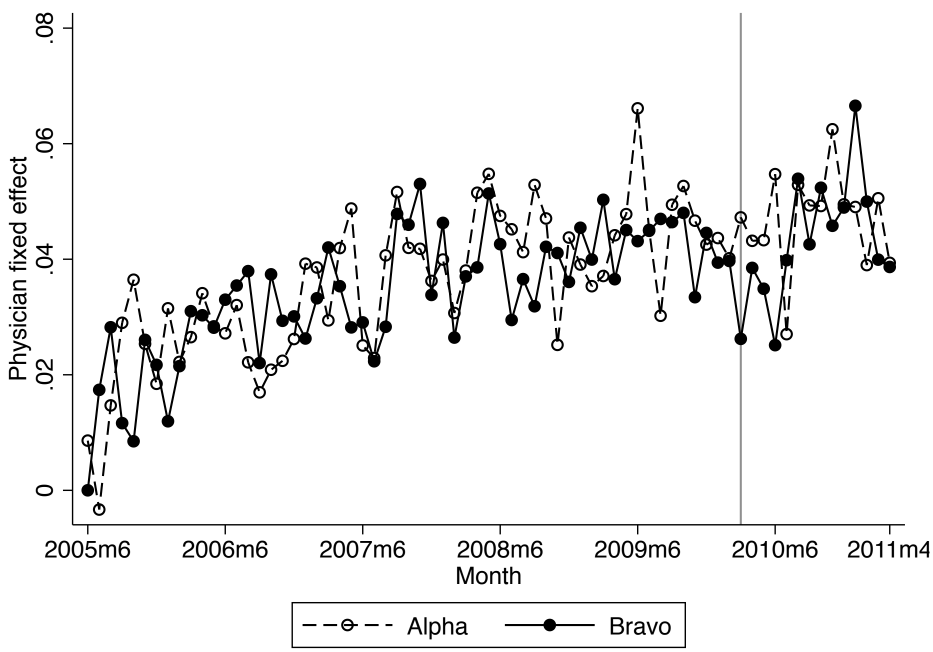
Note: This figure shows the average numbers of Elixhauser indices for patients in each pod, month, and year. Each Elixhauser index represents a clinical condition relevant for predicting patient outcomes, including congestive heart failure, diabetes, hypothyroidism, AIDS, metastatic cancer, and drug abuse (Elixhauser et al., 1998). Elixhauser indices are captured based on coding entered by health care providers in the medical record. Alpha pod averages are plotted with hollow circles; Bravo pod averages are plotted with solid circles. The vertical gray line indicates the month of the regime change of Bravo from a nurse-managed system to a self-managed system, in March 2010. Alpha was always self-managed.

Figure B-1.4: Average Predicted Log Length of Stay



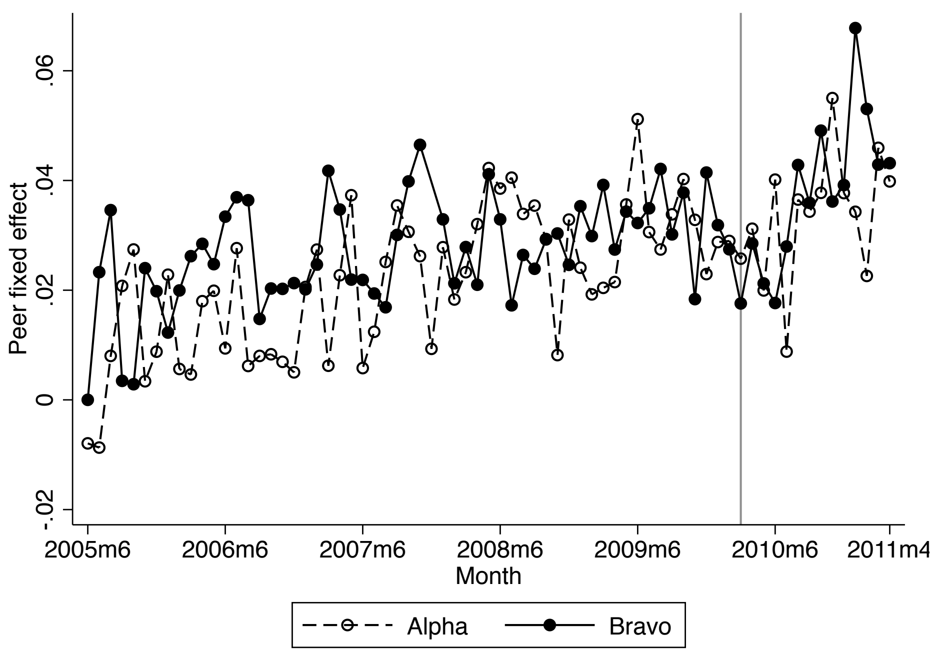
Note: This figure shows average predicted log length of stay for patients in each pod, month, and year. In the first step, Equation (B-1.1) is estimated for patients going to Alpha pod in 2005. Coefficients are then used to calculated predicted log lengths of stay for all patients. In the second step, averages for predicted log lengths of stay are computed for each pod, month, and year. Alpha pod averages are plotted with hollow circles; Bravo pod averages are plotted with solid circles. The vertical gray line indicates the month of the regime change of Bravo from a nurse-managed system to a self-managed system, in March 2010. Alpha was always self-managed.

Figure B-1.5: Average Physician Productivity



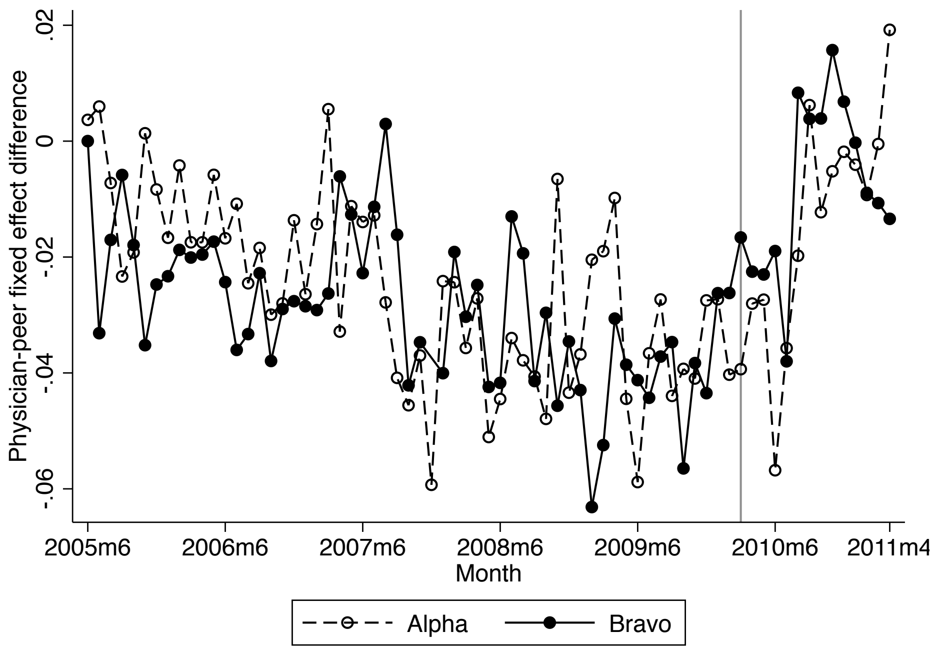
Note: This figure shows average productivity for physicians in each pod, month, and year. Physician productivity is first estimated as fixed effects in a regression of length of stay, described in Section B-1.2. In the second step, physician fixed effects are averaged for each pod, month, and year. Alpha pod averages are plotted with hollow circles; Bravo pod averages are plotted with solid circles. The vertical gray line indicates the month of the regime change of Bravo from a nurse-managed system to a self-managed system, in March 2010. Alpha was always self-managed.

Figure B-1.6: Average Peer Productivity



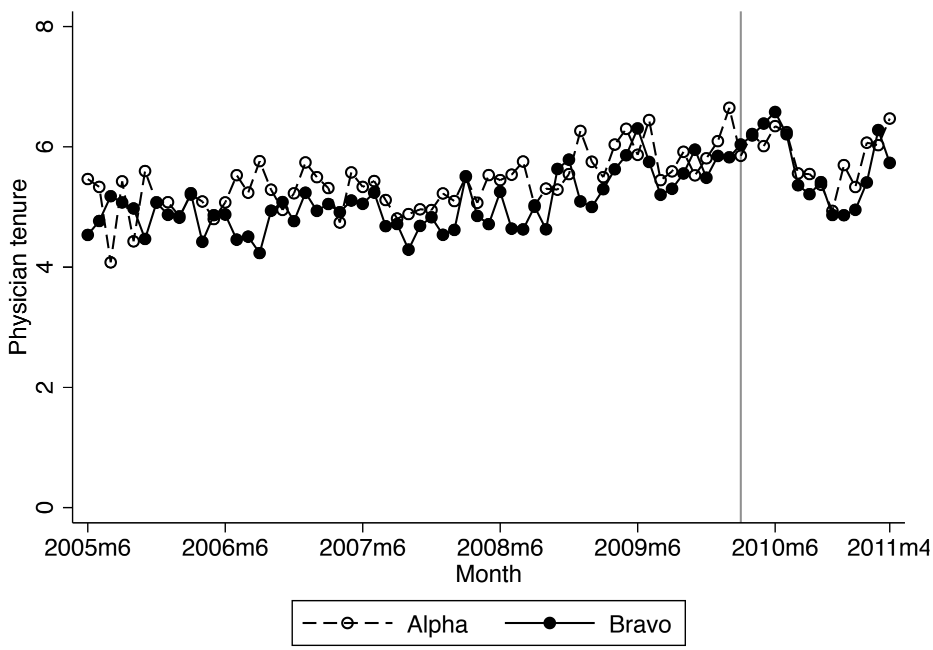
Note: This figure shows average productivity for peers (when present) in each pod, month, and year. Physician productivity is first estimated as fixed effects in a regression of length of stay, described in Section B-1.2. In the second step, fixed effects for peers are averaged for each pod, month, and year. Alpha pod averages are plotted with hollow circles; Bravo pod averages are plotted with solid circles. The vertical gray line indicates the month of the regime change of Bravo from a nurse-managed system to a self-managed system, in March 2010. Alpha was always self-managed.

Figure B-1.7: Average Physician-peer Productivity Differences



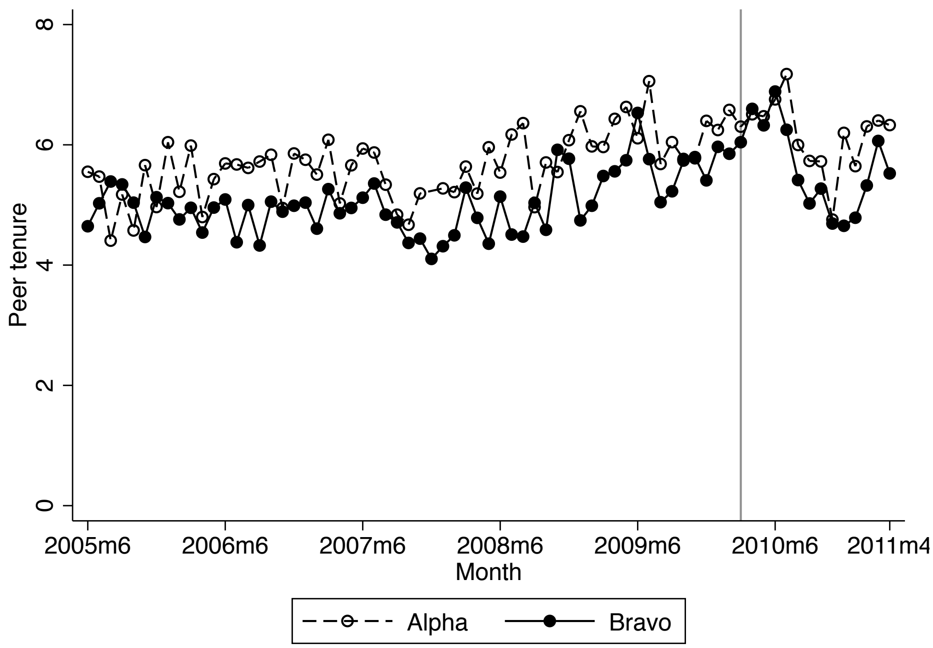
Note: This figure shows average differences in productivity between physicians and peers (when peers are present) in each pod, month, and year. Physician productivity is first estimated as fixed effects in a regression of length of stay, described in Section B-1.2. In the second step, differences in fixed effects between physicians and peers are averaged for each pod, month, and year. Alpha pod averages are plotted with hollow circles; Bravo pod averages are plotted with solid circles. The vertical gray line indicates the month of the regime change of Bravo from a nurse-managed system to a self-managed system, in March 2010. Alpha was always self-managed.

Figure B-1.8: Average Physician Tenure



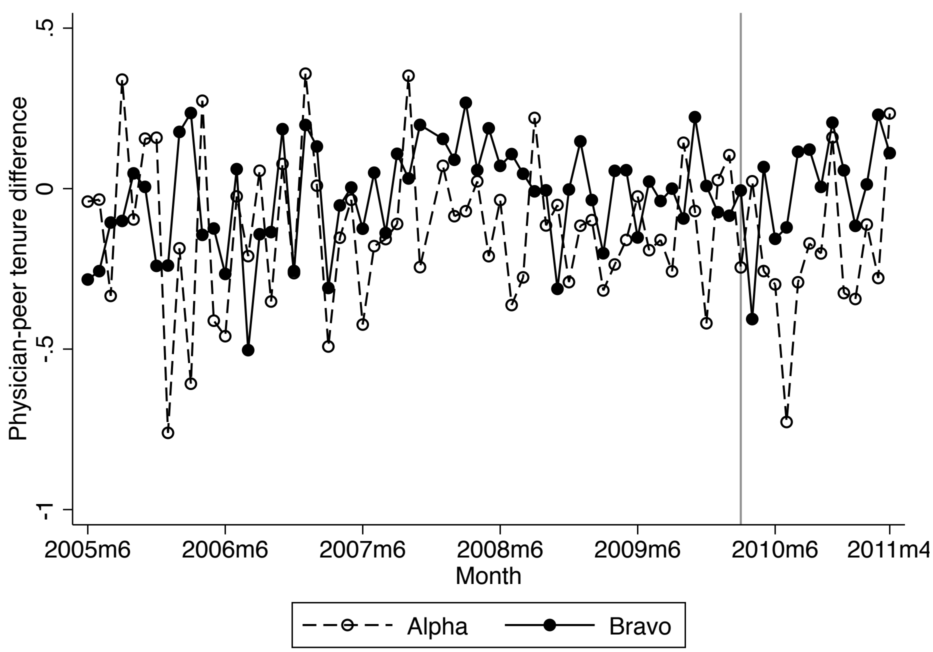
Note: This figure shows average tenure for physicians in each pod, month, and year. Physician tenure is measured by the difference between the patient visit and the physician date of hire. Alpha pod averages are plotted with hollow circles; Bravo pod averages are plotted with solid circles. The vertical gray line indicates the month of the regime change of Bravo from a nurse-managed system to a self-managed system, in March 2010. Alpha was always self-managed.

Figure B-1.9: Average Peer Tenure



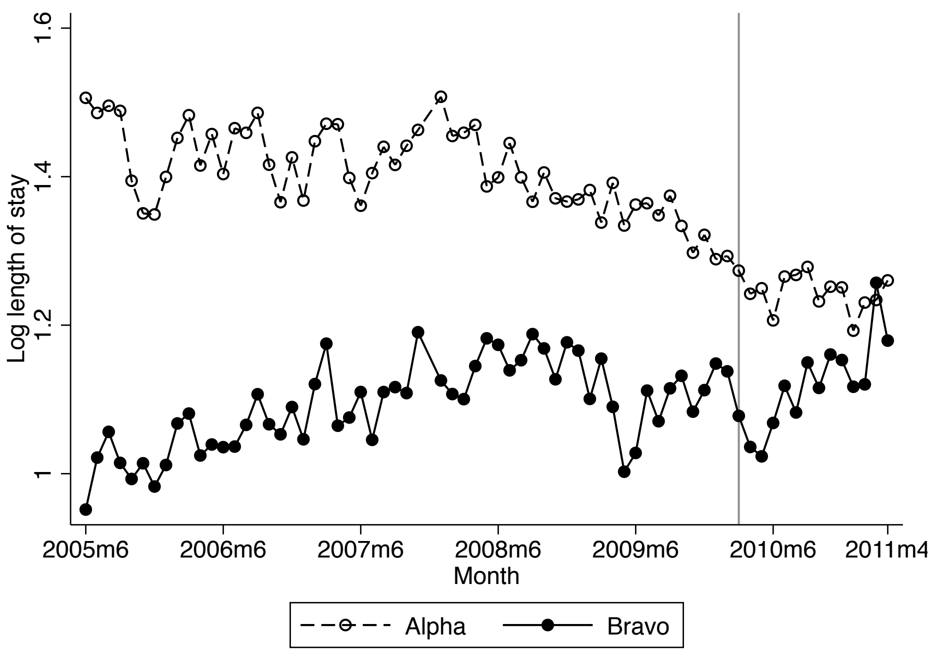
Note: This figure shows average tenure for peers (when present) in each pod, month, and year. Physician tenure is measured by the difference between the patient visit and the physician date of hire. Alpha pod averages are plotted with hollow circles; Bravo pod averages are plotted with solid circles. The vertical gray line indicates the month of the regime change of Bravo from a nurse-managed system to a self-managed system, in March 2010. Alpha was always self-managed.

Figure B-1.10: Average Physician-peer Differences in Tenure



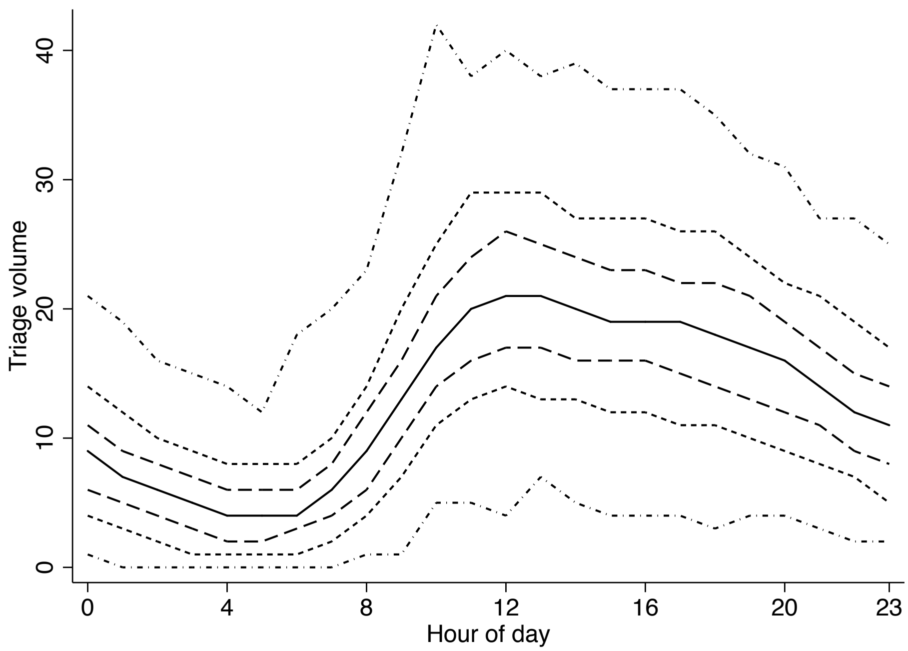
Note: This figure shows average differences in tenure between physicians and peers (when peers are present) in each pod, month, and year. Physician tenure is measured by the difference between the patient visit and the physician date of hire. Alpha pod averages are plotted with hollow circles; Bravo pod averages are plotted with solid circles. The vertical gray line indicates the month of the regime change of Bravo from a nurse-managed system to a self-managed system, in March 2010. Alpha was always self-managed.

Figure B-1.11: Unadjusted Average Log Length of Stay by Pod and Month-year



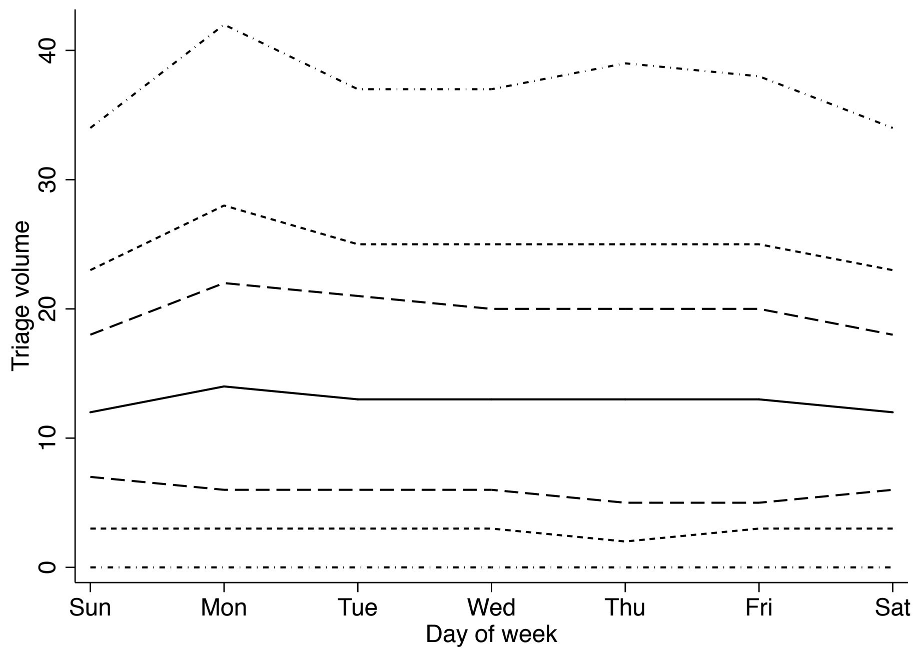
Note: This figure shows average log length of stay for each pod, month, and year, where the average is normalized to 0 for Bravo pod at the beginning of the panel. No patient characteristics or provider identities are controlled for, unlike Figure 2 in the main paper. Alpha pod averages are plotted with hollow circles; Bravo pod averages are plotted with solid circles. The vertical gray line indicates the month of the regime change of Bravo from a nurse-managed system to a self-managed system, in March 2010. Alpha was always self-managed.

Figure B-5.1: Overall Patient Volume by Hour



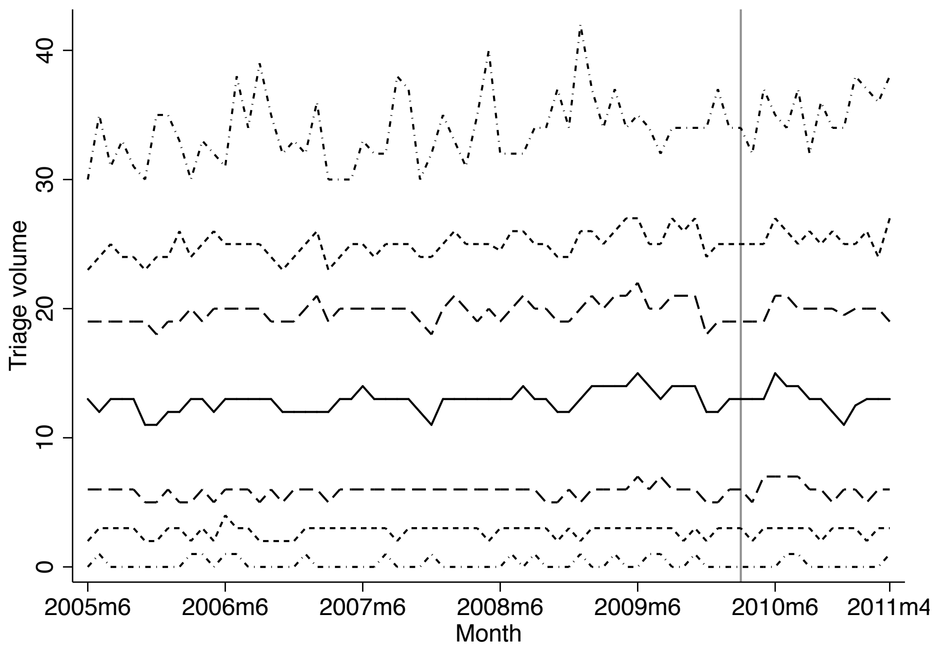
Note: This figure shows plots of overall patient volume by hour of the day. Patient volume is defined as the number of patients arriving to triage (the waiting room) in each hour. Summary statistics are then generated across these hourly observations for each hour of the day. The solid line plots median patient volume; the long-dashed lines plot 20th and 80th percentiles; the short-dashed lines plot the 5th and 95th percentiles; the dash-dotted lines plot minimum and maximum values.

Figure B-5.2: Overall Patient Volume by Day of Week



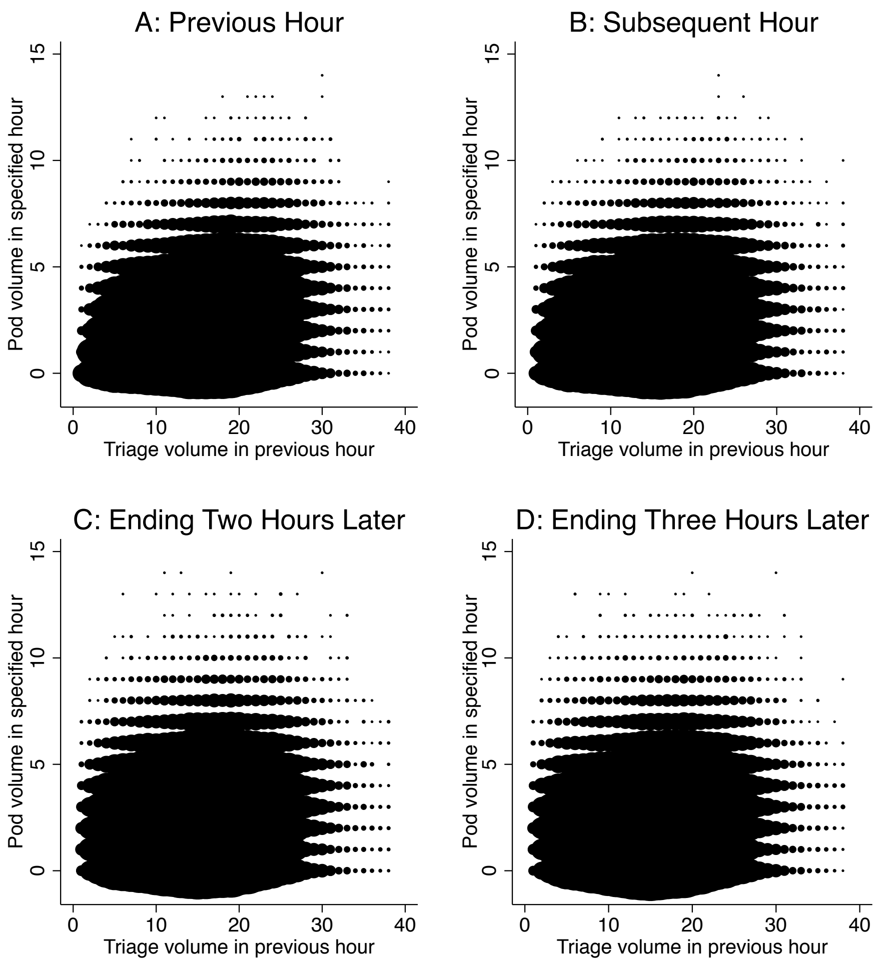
Note: This figure shows plots of overall patient volume by day of the week. Patient volume is defined as the number of patients arriving to triage (the waiting room) in each hour. Summary statistics are then generated across these hourly observations for each day of the week. The solid line plots median patient volume; the long-dashed lines plot 20th and 80th percentiles; the short-dashed lines plot the 5th and 95th percentiles; the dash-dotted lines plot minimum and maximum values.

Figure B-5.3: Overall Patient Volume by Month and Year



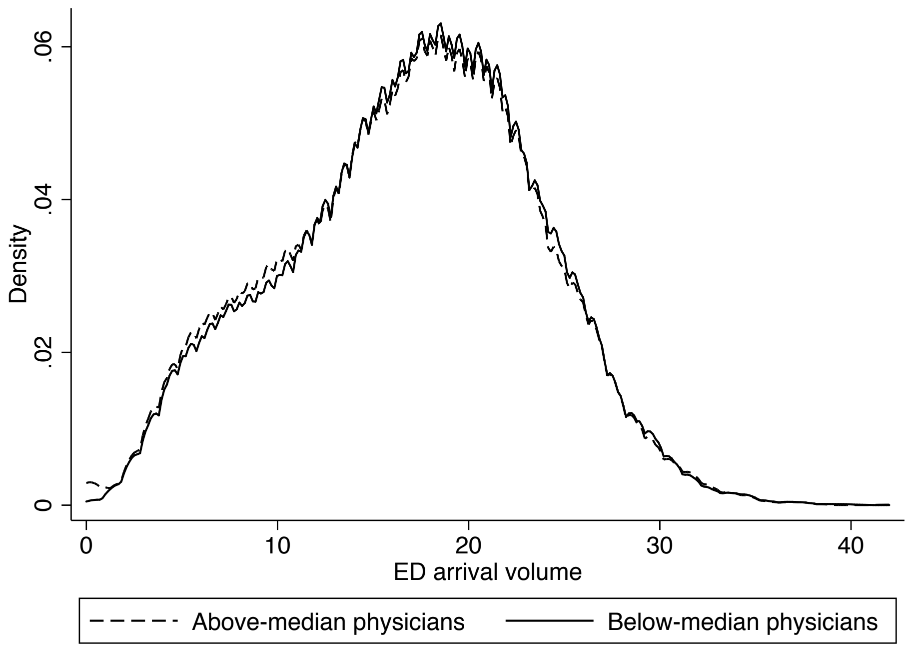
Note: This figure shows plots of overall patient volume by month and year. Patient volume is defined as the number of patients arriving to triage (the waiting room) in each hour. Summary statistics are then generated across these hourly observations for each month-year interaction. The solid line plots median patient volume; the long-dashed lines plot 20th and 80th percentiles; the short-dashed lines plot the 5th and 95th percentiles; the dash-dotted lines plot minimum and maximum values.

Figure B-5.4: Overall and Pod-specific Patient Volume



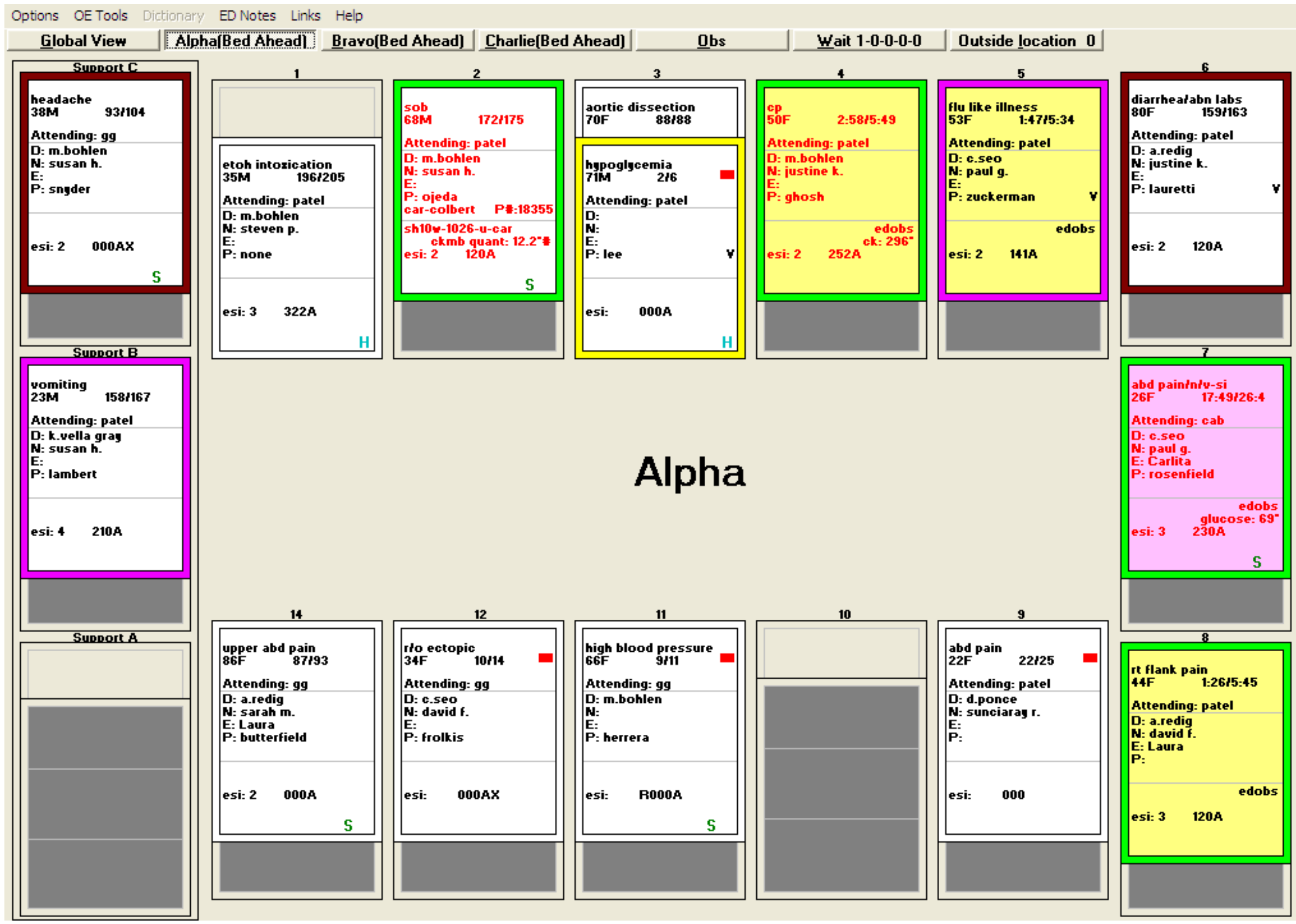
Note: This figure shows plots of overall and pod-specific patient volume. Circle sizes are proportional to the number of observations matching the specified overall and pod volumes. In all panels, overall volume is the number of patients arriving in triage in the hour prior to an index patient's pod arrival. Depending on the panel, pod-specific patient volume is shown for a specified hour, ranging from the same hour prior to an index patient's pod arrival to the hour ending three hours after the patient's pod arrival. The median time from triage to pod is about 30 minutes, while the median time from pod to discharge order is about 3.4 hours. Correlation coefficients are 0.21, 0.16, 0.13, and 0.12 for Panels A, B, C, and D, respectively.

Figure B-6.1: Density of Patient Volume to ED for High- and Low-Productivity Physicians



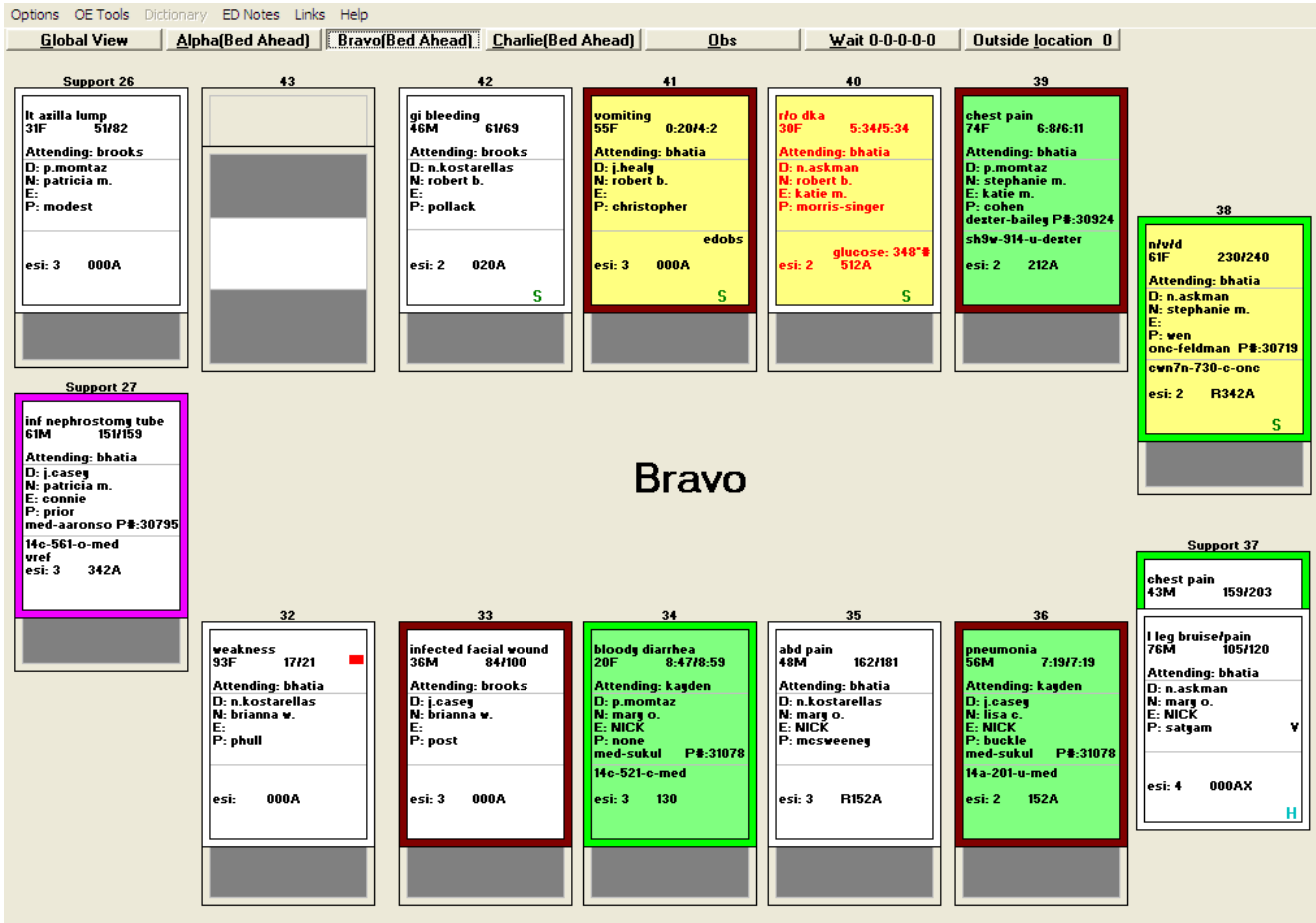
Note: This figure shows a kernel density plot of the ED arrival volume (number of patients arriving at the ED) during each hour while physicians of above- and below-median productivity are working. Dashed and solid lines show the densities for above- and below-median physicians, respectively. Physician productivity is estimated by fixed effects in a regression of length of stay, controlling for all possible interactions of team members (physician assistant or resident and nurse), coworker, pod location; patient demographics (age, sex, emergency severity index, Elixhauser comorbidities); ED arrival volume; and time dummies (month-year combination, day of the week, and hour of the day). The average difference in productivity between physicians of above- and below-median productivity is 0.28, meaning that physicians with above-average productivity take 28% less time than those with below-average productivity.

Figure B-8.1: Computer Schematic of Alpha Pod



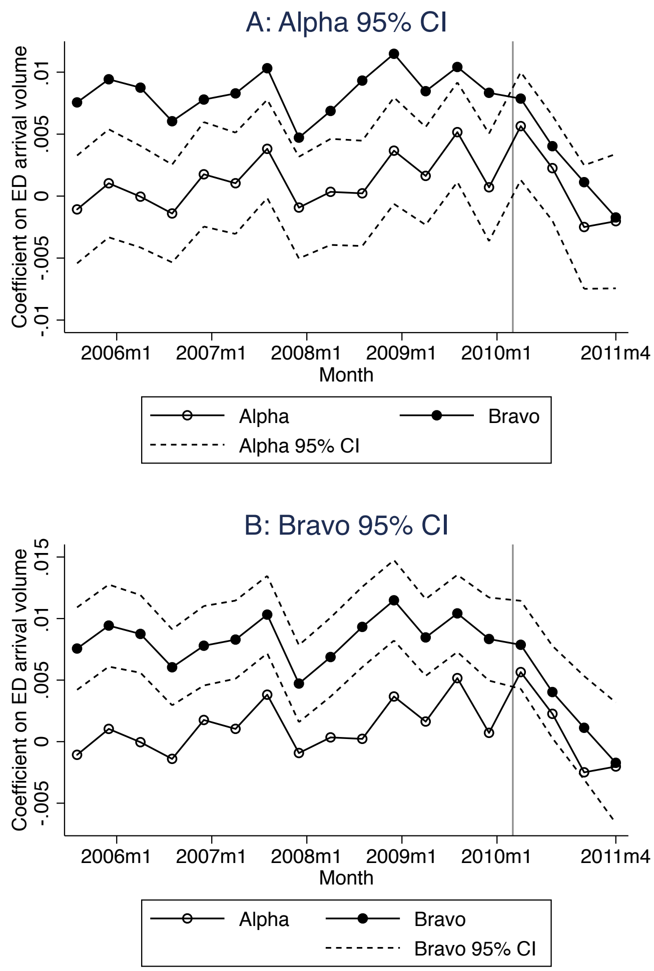
Note: This figure shows a computer screen layout of Alpha pod, which is both a spatial representation of the physical pod and the interface for physicians to select patients, examine the electronic medical record, and enter orders. Slots represent beds, with two beds per room, and filled beds are represented by slots with information. Colors represent various patient states, for example, whether an order needs to be taken off or whether the patient has been ordered for discharge. Identifying patient information has been removed here, but when displayed, such information includes patient last name, chief complaint, age, sex, physician, resident, nurse, emergency severity index, and minutes in ED (including triage) and in pod.

Figure B-8.2: Computer Schematic of Bravo Pod



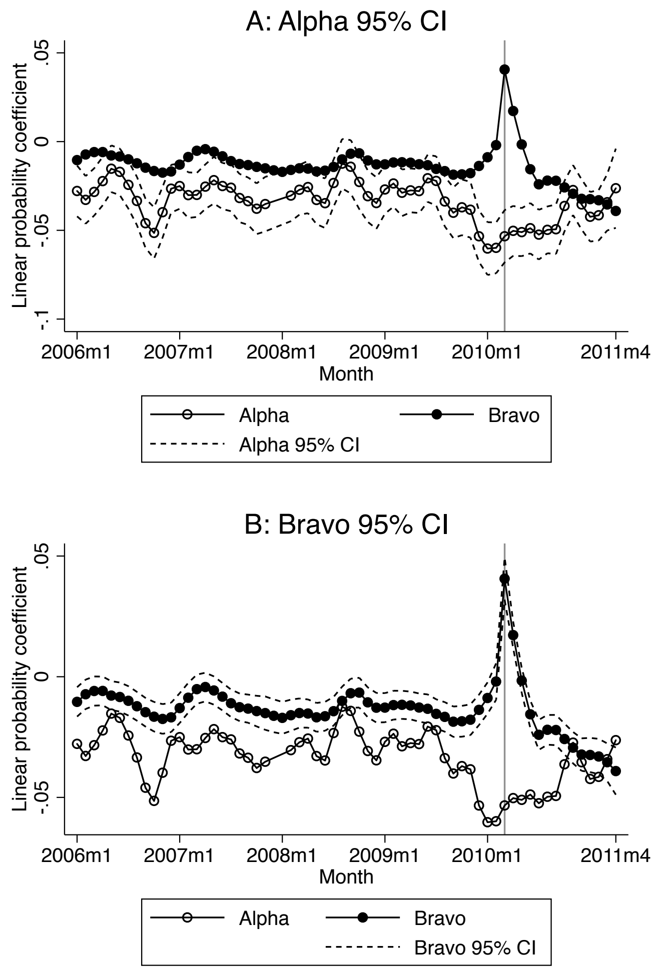
Note: This figure shows a computer screen layout of Bravo pod, which can be compared to Figure B-8.1 of Alpha pod. As in Figure B-8.1, this figure is both a spatial representation of the physical pod and the interface for physicians to select patients, examine the electronic medical record, and enter orders. Slots represent beds, with two beds per room, and filled beds are represented by slots with information. Prior to the regime change in March 2010, beds in Bravo was divided into two zones – Bravo-1 and Bravo-2 – each owned by a different physician. The location and numbering of beds were stable throughout the entire study period. Colors represent various patient states, for example, whether an order needs to be taken off or whether the patient has been ordered for discharge. Identifying patient information has been removed here, but when displayed, such information includes patient last name, chief complaint, age, sex, physician, resident, nurse, emergency severity index, and minutes in ED (including triage) and in pod.

Figure B-9.1: Event Study of Foot-dragging in Both Pods



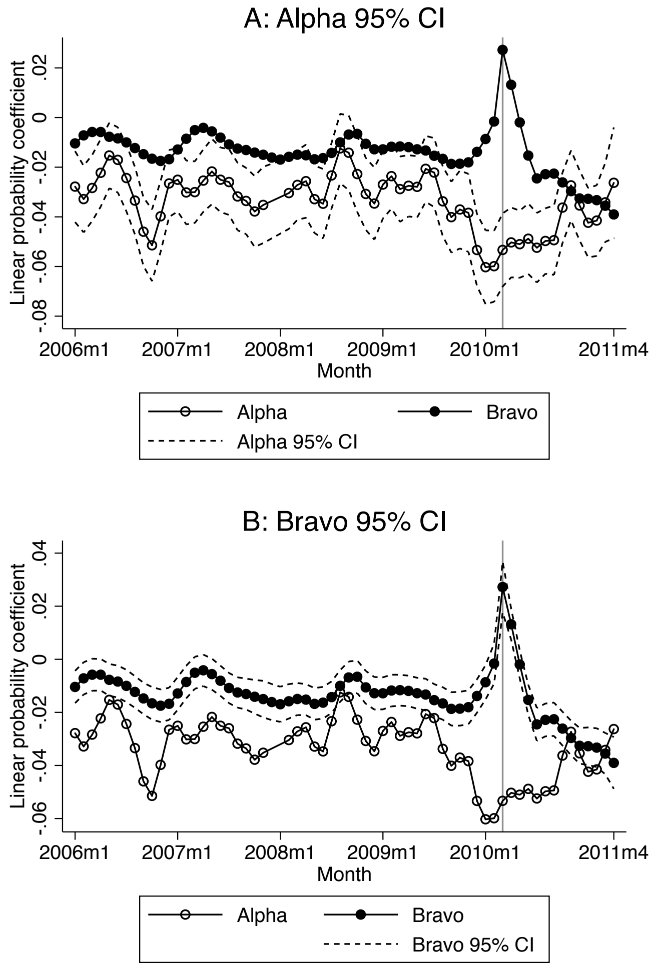
Note: This figure shows foot-dragging in both pods, as estimated by the log-length-of-stay coefficient for expected future work (measured by ED arrival volume, defined as the hourly rate of patients arriving at triage when the index patient arrives at the pod) interacted with pod identities and four-month interval dummies. These coefficients are plotted as an event study before and after the regime change of Bravo pod from a nurse-managed system to a self-managed system in March 2010, shown with a vertical gray line. Hollow and solid circles plot estimates for Alpha and Bravo pods, respectively. 95% confidence intervals, shown in dotted lines, are plotted in Panels A and B for Alpha and Bravo pods, respectively.

Figure B-9.2: New-patient Assignment Probability with Confidence Intervals



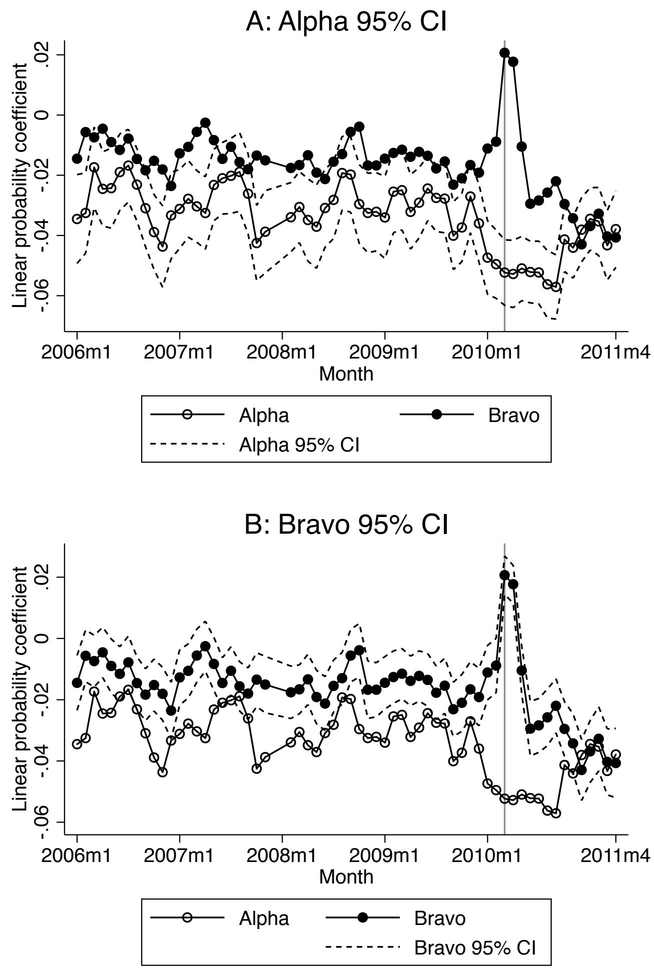
Note: This figure shows the new-patient assignment probability, as a function of relative censuses for physicians within each pod. The plotted coefficient estimates from Equation (7.1) represent the average effect on assignment probability of each additional patient on a physician's census relative to his peer's census. Hollow circles show coefficient estimates for Alpha pod, which was always self-managed. Solid circles show the coefficient estimates for Bravo pod, which switched to a self-managed system in March 2010, shown with a vertical gray line. Coefficients are estimated in a local linear regression using a triangular kernel with 90 days on each side, which may be truncated if sufficiently close to March 1, 2010 on either side. Panels A and B show confidence intervals for Alpha and Bravo coefficients, respectively.

Figure B-9.3: New-patient Assignment Probability, Alternative Specification



Note: This figure shows the new-patient assignment probability, as a function of relative censuses for physicians within each pod. Coefficients are estimated using Equation (B-9.1), slightly modified from Equation (7.1) by allowing for zone-based norms to continue in Bravo after the regime change. Coefficients represent the average effect on assignment probability of each additional patient on a physician's census relative to his peer's census. Hollow circles show coefficient estimates for Alpha pod, which was always self-managed. Solid circles show the coefficient estimates for Bravo pod, which switched to a self-managed system in March 2010, shown with a vertical gray line. Coefficients are estimated in a local linear regression using a triangular kernel with 90 days on each side, which may be truncated if sufficiently close to March 1, 2010. Panels A and B show confidence intervals for Alpha and Bravo coefficients, respectively.

Figure B-9.4: New-patient Assignment Probability, Kernel Regression



Note: This figure shows the new-patient assignment probability, as a function of relative censuses for physicians within each pod. The plotted coefficient estimates from Equation (7.1) represent the average effect on assignment probability of each additional patient on a physician's census relative to his peer's census. Hollow circles show coefficient estimates for Alpha pod, which was always self-managed. Solid circles show the coefficient estimates for Bravo pod, which switched to a self-managed system in March 2010, shown with a vertical gray line. Coefficients are estimated in a kernel regression using a triangular kernel with 45 days on each side, which may be truncated if sufficiently close to March 1, 2010. Panels A and B show confidence intervals for Alpha and Bravo coefficients, respectively.

Table B-1.1: Average Patient Characteristics Assigned to Alpha or Bravo Pod

Patient characteristic	Alpha	Bravo
Age	50.9 (19.8)	44.6 (18.9)
Emergency severity index	2.52 (0.697)	3.06 (0.785)
White	0.527 (0.499)	0.458 (0.498)
Black or African-American	0.217 (0.412)	0.240 (0.427)
Spanish speaking	0.082 (0.274)	0.109 (0.312)
Female and age $<$ 35 years	0.149 (0.357)	0.248 (0.432)

Note: This table reports average patient characteristics for patients being assigned to Alpha and Bravo pods. Alpha pod was always opened 24 hours, while Bravo pod always closed at night. The emergency severity index (ESI) ranges from 1 to 5, and a lower ESI represents a more severe patient. Standard deviations are reported in parentheses.

Table B-6.1: Average Patient Characteristics Available to Physicians with Preference for or against Patient Type

	Nurse-managed system		Self-managed system	
Patient characteristic	Physicians with preference for	Physicians with preference against	Physicians with preference for	Physicians with preference against
Age	44.4	44.3	50.6	50.2
	(18.8)	(18.8)	(19.7)	(19.7)
Emergency severity index	3.07	3.06	2.60	2.59
	(0.798)	(0.791)	(0.719)	(0.720)
White	0.461	0.455	0.532	0.529
	(0.498)	(0.498)	(0.499)	(0.499)
Black or African-American	0.245	0.247	0.227	0.228
	(0.430)	(0.431)	(0.419)	(0.419)
Spanish speaking	0.111	0.111	0.091	0.090
	(0.314)	(0.315)	(0.288)	(0.286)
Female and age $<$ 35 years	0.251	0.249	0.157	0.164
	(0.434)	(0.432)	(0.364)	(0.370)

Note: This table reports average patient characteristics available to physicians with a preference for or against a given patient type, while working in a nurse-managed or self-managed system. In the nurse-managed system, patients are assigned; on the self-managed system, patients are made available by the triage nurse. I construct measures of physician preference by (1) estimating a linear probability model of patient choice using observations in the self-managed system and including physician-specific coefficients on patient characteristics, and (2) selecting physicians with high or low coefficients as having a preference for or against each patient characteristic, respectively. Once I arrive at measures of physician preference for each patient characteristic, I describe the average for that patient characteristic for patients available to choose from for physicians with preferences for or against that characteristic. The emergency severity index (ESI) ranges from 1 to 5, and a lower ESI represents a more severe patient. Standard deviations are reported in parentheses.

Table B-7.1: Peer Effects on Log Length of Stay

	(1)	(2)	(3)	(4)	(5)	(6)
Peer effect	0.115***	0.087*	0.106***			
	(0.035)	(0.049)	(0.032)			
Peer effect, faster than median				0.142***	0.079	
				(0.047)	(0.069)	
Peer effect, slower than median				0.039	0.106	
				(0.098)	(0.124)	
Peer effect × self-managed			-0.053			
			(0.053)			
Peer effect, faster than median × self-managed						0.109*
						(0.062)
Peer effect, slower than median× self-managed						-0.081
						(0.115)
Peer effect, faster than median × nurse-managed						0.128***
						(0.045)
Peer effect, slower than median× nurse-managed						0.048
						(0.093)
Self-managed			-0.081***			-0.061**
			(0.029)			(0.024)
Sample	Nurse-managed	Self-managed	Pooled	Nurse-managed	Self-managed	Pooled
Number of observations	129,274	129,798	259,072	129,274	129,798	259,072
Adjusted R-squared	0.485	0.378	0.404	0.485	0.378	0.404
Sample log length of stay mean	1.006	1.263	1.135	1.006	1.263	1.135

Note: I estimate physician peer effects depending on the organizational structure. I first estimate fixed effects in a regression of productivity on physician identities, adjusting for all possible interactions with resident, nurse, peer, and pod location. I then use these fixed effects for observations in the mechanical system where there is another physician in the same pod present to estimate a regression of log length of stay with peer effects. Models (1) and (4) are estimated for observations from nurse-managed teams, models (2) and (5) are estimated for observations from self-managed teams, and models (3) and (6) estimated from pooled observations. Models (4) to (6) split peer effects for peers that are slower or faster than median. All regressions in both stages are adjusted for hour of the day, day of the week, month-year dummies, Elixhauser score dummies, emergency severity index, ED arrival volume, patient demographics, and patient triage time. All models are clustered for physician-resident-nurse trio identities.

Table B-7.2: Effect of Peer Relationships on Foot-dragging

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	Log length of stay			Log length of stay			Log length of stay		
Type of peer	Greater tenure than physician			Above-median productivity			$\geq$ 60 hours prior work together		
ED volume	0.008***	0.003***	0.009***	0.008***	0.004***	0.009***	0.006***	0.004***	0.007***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
ED volume × peer type	-0.004***	0.000	-0.003***	-0.002*	-0.001	-0.001	0.002	-0.002	0.002*
	(0.001)	(0.002)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Self-managed × ED volume			-0.008***			-0.007***			-0.005***
			(0.001)			(0.001)			(0.001)
Self-managed × ED volume × peer type			0.002			0.001			-0.004*
			(0.002)			(0.002)			(0.002)
Sample	Nurse-managed	Self-managed	Pooled	Nurse-managed	Self-managed	Pooled	Nurse-managed	Self-managed	Pooled
Observations	121,024	126,264	247,288	121,024	126,264	247,288	121,024	126,254	247,278
Adjusted R-squared	0.445	0.364	0.376	0.445	0.364	0.376	0.445	0.364	0.376
Sample mean log (hours) length of stay	0.936	1.251	1.097	0.926	1.179	1.064	0.926	1.179	1.064
Sample mean ED volume	17.30	15.08	16.16	17.12	14.22	15.53	17.12	14.22	15.53

Note: This table reports effect of expected future work, interacted with a peer type, as by Equation (B-7.3). Expected future work is measured by ED arrival volume ("ED volume" for brevity), defined as the number of patients arriving at ED triage during the hour prior to the index patient's arrival at the pod. All observations require the presence of a peer. Peers with greater tenure by at least 2 years, those with high productivity (faster-than-median fixed effect for lengths of stay), and familiar peers (peers with at least 60 hours of history, approximately the 75th percentile, working with the index physician) are considered. ED volume is demeaned. Coefficients for the direct effect of the peer type, self-managed (in pooled models), and self-managed × peer type (in pooled models) are omitted for brevity. All models control for time categories (month-year, day of the week, and hour of the day dummies), pod (when applicable), patient demographics (age, sex, race, and language), patient clinical information (Elixhauser comorbidity indices, emergency severity index), triage time, and physician-resident-nurse interactions. The nurse-managed, self-managed, and pooled samples had 121,024, 126,264, and 247,288 observations, respectively. All models are clustered by physician.
* significant at 10%; ** significant at 5%; *** significant at 1%.

Table B-8.1: Bed Locations

Pod or zone	Bed locations
Alpha	1, 1H, 2, 2H, 3, 3H, 4, 4H, 5, 5H, 6, 6H, 7, 7H, 8, 8H, 9, 9H, 10, 10H, 11, 11H, 12, 12H, 14, 14H, A Hall-A, A Hall-AH, A Hall-B, A Hall-BH, A Hall-C, A Hall-CH, A-HALL
Bravo-1	25, 25H, 26, 26H, 27, 27H, 28, 29, 30, 31, 32, 32H, 43, 43H, E Hall-A, E Hall-AH, E Hall-B, E Hall-BH
Bravo-2	33, 33H, 34, 34H, 35, 35H, 36, 36H, 37A, 37AH, 37B, 37BH, 38, 38H, 39, 39H, 40, 40H, 41, 41H, 42, 42H, B-HALL

Note: This table details bed locations for beds considered in each pod or zone. Alpha and Bravo pods are separate locations in the ED. Prior to March 2010, Bravo pod was divided into two non-overlapping zones – Bravo-1 and Bravo-2 – and physicians in Bravo were scheduled to care for patients in one of these two zones. Alpha pod is shown in Figure 2 in the main paper; Bravo pod is shown in Figure B-8.2. Some of the bed locations are visible in these figures, while others are not annotated, as they may represent a "virtual bed," such as the hallway space in front of a bed, for example. All bed locations correspond to unique electronic designations for a physical space in the ED. There was no change in the physical arrangement of non-virtual beds in either pod throughout the sample period.

Table B-8.2: Alpha Bed Visits

Bed	Dates observed	Number of visits
1	6/1/2005-4/30/2011	15,560
1H	1/22/2008-4/30/2011	5,380
2	6/1/2005-4/30/2011	15,945
2H	1/22/2008-4/30/2011	4,885
3	6/1/2005-4/30/2011	13,385
3H	1/22/2008-4/30/2011	4,210
4	6/1/2005-4/30/2011	13,103
4H	1/22/2008-4/30/2011	4,642
5	6/1/2005-4/30/2011	12,722
5H	1/22/2008-4/30/2011	3,638
6	6/1/2005-4/30/2011	11,995
6H	1/21/2008-4/30/2011	1,642
7	6/1/2005-4/30/2011	10,634
7H	1/21/2008-4/29/2011	1,279
8	6/1/2005-4/30/2011	10,423
8H	1/21/2008-4/28/2011	1,137
9	6/1/2005-4/30/2011	11,819
9H	1/21/2008-4/30/2011	4,493
10	6/1/2005-4/30/2011	11,576
10H	1/21/2008-4/30/2011	2,220
11	6/1/2005-4/30/2011	11,625
11H	1/21/2008-4/30/2011	1,976
12	6/1/2005-4/30/2011	12,279
12H	1/22/2008-4/29/2011	3,112
14	6/1/2005-4/30/2011	12,769
14H	1/21/2008-4/30/2011	5,358
A Hall-A	1/21/2008-4/20/2011	6,354
A Hall-AH	1/21/2008-4/17/2011	1,692
A Hall-B	1/21/2008-4/20/2011	5,652
A Hall-BH	1/21/2008-4/15/2011	1,429
A Hall-C	1/21/2008-4/19/2011	6,432
A Hall-CH	1/21/2008-4/19/2011	1,735
A-HALL	6/1/2005-1/21/2008	24800

Note: This table shows the number of visits for which a patient is recorded to spend some time in a given bed in Alpha pod. Date ranges are also given for the first and last of these visits. Alpha pod is shown in Figure 2 in the main paper. Some of the bed locations are visible in this figures, while others are not annotated, as they may represent a "virtual bed," such as the hallway space in front of a bed, for example. Hallway beds are designated with an H and sometimes a number signifying the closest numbered room. All bed locations correspond to unique electronic designations for a physical space in the ED. There was no change in the physical arrangement of non-virtual beds throughout the sample period.

Table B-8.3: Bravo-1 Bed Visits

Bed	Dates observed	Number of visits
25	6/1/2005-2/15/2011	7,415
25H	1/22/2008-4/19/2011	1,229
26	6/1/2005-4/30/2011	9,043
26H	1/22/2008-4/28/2011	1,642
27	6/1/2005-4/30/2011	9,812
27H	1/22/2008-4/30/2011	1,474
28	6/1/2005-2/7/2010	5,719
29	6/1/2005-2/3/2010	7,051
30	6/1/2005-3/29/2010	8,593
31	6/1/2005-1/26/2010	9,808
32	6/1/2005-4/30/2011	9,343
32H	1/22/2008-4/30/2011	2,278
43	6/1/2005-4/30/2011	9,970
43H	1/22/2008-4/30/2011	2,073
E Hall-A	1/24/2008-4/5/2010	1,936
E Hall-AH	1/22/2008-4/5/2010	1,768
E Hall-B	1/24/2008-4/5/2010	1,136
E Hall-BH	1/22/2008-4/5/2010	845

Note: This table shows the number of visits for which a patient is recorded to spend some time in a given bed in Bravo-1 zone within pod. Date ranges are also given for the first and last of these visits. Bravo-1 and Bravo-2 designate adjoining but non-overlapping zones within Bravo, each containing beds that are owned by a physician working in shifts labeled with "Bravo-1" or "Bravo-2," respectively. Bravo pod is shown in Figure B-8.2. Some of the bed locations are visible in this figures, while others are not annotated, as they may represent a "virtual bed," such as the hallway space in front of a bed, for example. Hallway beds are designated with an H and sometimes a number signifying the closest numbered room. Beds 28 to 31 are spaces without walls but separated by curtains. All bed locations correspond to unique electronic designations for a physical space in the ED. There was no change in the physical arrangement of non-virtual beds throughout the sample period.

Table B-8.4: Bravo-2 Bed Visits

Bed	Dates observed	Number of visits
33	6/1/2005-4/30/2011	11,613
33H	1/22/2008-4/29/2011	3,982
34	6/1/2005-4/30/2011	12,186
34H	1/22/2008-4/29/2011	4,504
35	6/1/2005-4/30/2011	11,186
35H	1/22/2008-4/30/2011	2,323
36	6/1/2005-4/30/2011	10,669
36H	1/22/2008-4/30/2011	1,905
37A	6/1/2005-4/19/2011	11,199
37A H	1/23/2008-4/19/2011	1,682
37B	6/1/2005-4/19/2011	10,357
37B H	1/22/2008-4/18/2011	1,036
38	6/1/2005-4/30/2011	11,492
38H	1/22/2008-4/29/2011	1,956
39	6/1/2005-4/30/2011	13,067
39H	1/22/2008-4/30/2011	5,715
40	6/1/2005-4/30/2011	12,772
40H	1/22/2008-4/28/2011	2,678
41	6/1/2005-4/30/2011	13,047
41H	1/22/2008-4/30/2011	2,726
42	6/1/2005-4/30/2011	12,915
42H	1/22/2008-4/30/2011	4,279
B-HALL	6/1/2005-1/21/2008	14,771

Note: This table shows the number of visits for which a patient is recorded to spend some time in a given bed in Bravo-2 zone within pod. Date ranges are also given for the first and last of these visits. Bravo-1 and Bravo-2 designate adjoining but non-overlapping zones within Bravo, each containing beds that are owned by a physician working in shifts labeled with "Bravo-1" or "Bravo-2," respectively. Bravo pod is shown in Figure B-8.2. Some of the bed locations are visible in this figures, while others are not annotated, as they may represent a "virtual bed," such as the hallway space in front of a bed, for example. Hallway beds are designated with an H and sometimes a number signifying the closest numbered room. All bed locations correspond to unique electronic designations for a physical space in the ED. There was no change in the physical arrangement of non-virtual beds throughout the sample period.

Table B-9.1: Overall Effect of Self-managed System on Other Outcomes

	(1)	(2)	(3)	(4)	(5)	(6)
	30-day mortality	Hospital admissions	14-day return visits	Relative Value Units (RVUs)	Log total costs	Log time to first order
Self-managed system						
Model 1	0.0041	0.0114	-0.0117*	0.009	0.018	-0.001
	(0.0032)	(0.0091)	(0.0067)	(0.028)	(0.032)	(0.024)
Model 2	0.0031	0.0049	-0.0114*	-0.003	0.004	-0.002
	(0.0031)	(0.0088)	(0.0067)	(0.0028)	(0.030)	(0.024)
Model 3	-0.0002	-0.0010	-0.0121*	-0.014	-0.021	-0.003
	(0.0031)	(0.0086)	(0.0067)	(0.028)	(0.030)	(0.024)
Model 4	0.0002	0.0004	-0.0127*	-0.013	-0.016	-0.025
	(0.0031)	(0.0086)	(0.0067)	(0.028)	(0.030)	(0.024)
Model 5	0.0048	0.0020	-0.0129*	0.023	0.056	-0.102***
	(0.0038)	(0.0107)	(0.0084)	(0.037)	(0.037)	(0.030)
Number of observations	317,199	317,199	317,199	255,516	284,965	307,885
Sample mean outcome	0.020	0.272	0.067	2.701	6.724	-0.623

Note: This table extends Table 3, which reports the effect of the self-managed system on outcomes other than length of stay, by presenting results for each of the five models in Table 2. Recall that Table 3 presents only results for Model 3. As in Table 2, models increasingly include controls to Equation (4.2). Relative Value Units (RVUs) represent intensity of care and are directly scalable to dollar amounts of clinical revenue. Total costs include direct costs for the entire visit, which may include hospital admission. Time to first order is the time between patient arrival at the pod and the first physician order, measured in log hours. All models control for ED patient arrival volume, time categories (hour of the day, day of the week, and month-year dummies), pod, and physician-resident-nurse interactions. All models are clustered by physician. * significant at 10%; ** significant at 5%; *** significant at 1%.

Table B-9.2: Effect of Expected Future Work on Other Outcome and Process Measures

	Outcomes				Process Measures			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	30-day mortality	Hospital Admissions	14-day return visits	Log total costs	Total order count	Diagnostic order count	Medication order count	CT or MRI ordered
ED volume	-0.0001	-0.0006*	-0.0002	-0.0022*	-0.0117*	-0.0041	-0.0050*	0.0002
	(0.0001)	(0.0004)	(0.0002)	(0.0012)	(0.0071)	(0.0042)	(0.0029)	(0.0003)
ED volume × self-managed	0.0002	-0.0006	0.0001	-0.0011	-0.0094	-0.0027	-0.0024	-0.0002
	(0.0003)	(0.0004)	(0.0003)	(0.0027)	(0.0101)	(0.0060)	(0.0060)	(0.0005)
Self-managed	-0.001	0.004	-0.011	-0.0062	-0.526*	-0.060	-0.109	-0.021
	(0.006)	(0.012)	(0.009)	(0.0642)	(0.280)	(0.167)	(0.164)	(0.013)
Adjusted R-squared	0.338	0.461	-0.038	0.526	0.548	0.489	0.402	0.295
Sample mean outcome	0.019	0.265	0.066	6.684	13.267	5.337	2.688	0.246

Note: This table shows the effect of expected future work on other outcomes and process measures, estimated by Equation (5.2). Log total costs include all direct costs incurred from the encounter, including any from admissions. Expected future work is measured by ED arrival volume ("ED volume" for brevity), defined as the number of patients arriving at ED triage during the hour prior to the index patient's arrival at the pod. Models do not include pod-level prior patient volume; results are unchanged when including this. All models control for time categories (month-year, day of the week, and hour of the day dummies), pod, patient demographics (age, sex, race, and language), patient clinical information (Elixhauser comorbidity indices, emergency severity index), triage time, and physician-resident-nurse interactions. Results are insensitive to controlling for pod-level patient volume. All models are clustered by physician. All models have 289,132 observations, except for model (4), which has 269,905 observations. The sample mean patient volume is 15.53 for all models, except for the model (4), for which it is 15.48. * significant at 10%; ** significant at 5%; *** significant at 1%.

Table B-9.3: Effect of Peer Presence on Foot-dragging, ED Volume Measured per Physician

	(1)	(2)	(3)
	Log length of stay		
Sample	Nurse-managed	Self-managed	Pooled
ED volume per physician present	0.026***	0.001	
	(0.003)	(0.002)	
× no peer present	0.037***	0.022***	0.039***
	(0.006)	(0.006)	(0.003)
× peer present, nurse-managed			0.032***
			(0.002)
× peer present, self-managed			-0.002
			(0.004)
× only physician in ED			0.003
			(0.002)
Number of observations	130,581	165,610	296,113
Adjusted R-squared	0.441	0.353	0.365
Sample mean log length of stay (log hours)	0.926	1.168	1.061
Sample mean ED volume per physician	4.702	4.139	4.388

Note: This table is similar to Table 5, except that it uses ED arrival volume per physician present as the measure of expected future work. Notes are otherwise the same as for Table 5. * significant at 10%; ** significant at 5%; *** significant at 1%.